

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 16:25:44 ; Search time 27.1642 Seconds
(without alignments)
540.877 Million cell updates/sec

Title: US-09-423-100-5

Perfect score: 294

Sequence: 1 FVNQHLCGSHLVEALYLVCG.....IVEQCCTSICSLYQLENYCN 52

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:
1: geneseqp1980s:
2: geneseqp1990s:
3: geneseqp2000s:
4: geneseqp2001s:
5: geneseqp2002s:
6: geneseqp2003as:
7: geneseqp2003bs:
8: geneseqp2004s:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query				Description
		Match	Length	DB	ID	
1	294	100.0	52	2	AAY42859	Aay42859 Human ins
2	294	100.0	56	2	AAR68901	Aar68901 Human pro
3	294	100.0	56	2	AAR78665	Aar78665 Proinsuli
4	294	100.0	63	2	AAR68900	Aar68900 Human pro
5	294	100.0	96	2	AAR68899	Aar68899 Human pro
6	294	100.0	96	2	AAR78662	Aar78662 Fusion pr
7	294	100.0	107	2	AAY42860	Aay42860 hGH-mini-
8	294	100.0	116	2	AAR98897	Aar98897 SOD-proin
9	294	100.0	137	2	AAR71692	Aar71692 Mating fa

10	294	100.0	145	2	AAR71694	Aar71694 Mating fa
11	294	100.0	146	2	AAR71695	Aar71695 Mating fa
12	294	100.0	150	2	AAY42861	Aay42861 Chimeric
13	291	99.0	57	2	AAR04582	Aar04582 Proinsuli
14	287	97.6	52	2	AAR11899	Aar11899 Example o
15	283.5	96.4	53	2	AAR65883	Aar65883 Di-Arg- (B
16	283.5	96.4	53	2	AAW18007	Aaw18007 Ins1 doub
17	283.5	96.4	160	2	AAR79056	Aar79056 Glycosylp
18	282.5	96.1	117	2	AAR98896	Aar98896 SOD-proin
19	281.5	95.7	60	1	AAP20002	Aap20002 Human pro
20	278.5	94.7	51	3	AAB12771	Aab12771 Human mat
21	278.5	94.7	51	3	AAB12774	Aab12774 Human mat
22	278.5	94.7	51	5	AAM48219	Aam48219 Human ins
23	278.5	94.7	53	2	AAR96046	Aar96046 Modified
24	278.5	94.7	55	2	AAR96050	Aar96050 Modified
25	278.5	94.7	63	2	AAW06807	Aaw06807 Single ch
26	278.5	94.7	66	2	AAW06809	Aaw06809 Met-Arg-M
27	278.5	94.7	87	2	AAW06810	Aaw06810 Single ch
28	278.5	94.7	138	2	AAR87086	Aar87086 pKV142 mo
29	277.5	94.4	53	1	AAP60132	Aap60132 Sequence
30	277.5	94.4	53	2	AAR65882	Aar65882 Gly-A21-d
31	277	94.2	65	2	AAW47365	Aaw47365 Preproins
32	276.5	94.0	67	2	AAW17998	Aaw17998 Ins2 doub
33	276.5	94.0	67	2	AAW17999	Aaw17999 Ins3 doub
34	275.5	93.7	53	2	AAR11898	Aar11898 Example o
35	275.5	93.7	53	2	AAR87085	Aar87085 pAK679 mo
36	275.5	93.7	53	2	AAY28341	Aay28341 Modified
37	275.5	93.7	53	4	AAB30704	Aab30704 Amino aci
38	275.5	93.7	55	1	AAP71019	Aap71019 Sequence
39	275.5	93.7	58	2	AAR96047	Aar96047 Modified
40	275.5	93.7	59	2	AAR96048	Aar96048 Modified
41	275.5	93.7	61	2	AAR96049	Aar96049 Modified
42	275.5	93.7	65	2	AAR88188	Aar88188 N-termina
43	275.5	93.7	89	2	AAR88179	Aar88179 Signal pe
44	275.5	93.7	91	2	AAR25584	Aar25584 Insulin p
45	275.5	93.7	91	2	AAR88181	Aar88181 YAP3 sign

ALIGNMENTS

RESULT 1
AAY42859
ID AAY42859 standard; protein; 52 AA.
XX
AC AAY42859;
XX
DT 19-JAN-2000 (first entry)
XX
DE Human insulin precursor, SEQ ID 5.
XX
KW Insulin; precursor; growth hormone; chaperone; intramolecular; folding;
KW conformation; chimeric protein; cleavable; recombinant; production;
KW yield.
XX
OS Homo sapiens.
XX

PN WO9950302-A1.
XX
PD 07-OCT-1999.
XX
PF 31-MAR-1998; 98WO-CN000052.
XX
PR 31-MAR-1998; 98WO-CN000052.
XX
PA (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.
XX
PI Gan Z;
XX
DR WPI; 1999-610839/52.
XX
PT New chimeric proteins containing human growth hormone fragment, used
PT particularly for the production of human insulin.
XX
PS Claim 12; Page 29-30; 46pp; English.
XX
CC This sequence represents a human insulin precursor comprising insulin A
CC and B chains. This insulin precursor is a component of the chimeric
CC proteins hGH-mini-proinsulin (AAY42860) and the chimeric protein given in
CC AAY42861. These chimeric proteins additionally contain an N-terminal
CC fragment of human growth hormone (hGH) and a cleavable peptide linker
CC (AAY42857). The hGH portion of the chimeric protein acts as an
CC intramolecular chaperone (IMC) for the insulin precursor, enabling it to
CC fold correctly. The cleavable peptide linker has a C-terminal Arg residue
CC which enables the hGH portion of the chimeric protein to be removed after
CC folding has taken place. Production of recombinant human insulin via an
CC hGH-proinsulin chimeric protein can provide human insulin with correctly
CC linked cysteine bridges with fewer necessary procedural steps, and hence
CC resulting in a higher yield of human insulin. The IMC sequences not only
CC protect insulin sequences from intracellular degradation by a
CC microorganism host, but also promote the folding of the fused insulin
CC precursor, facilitate the solubility of the fusion protein and decrease
CC the intermolecular interactions among the fusion proteins, thus allowing
CC folding of the fused insulin precursor at commercially useful high
CC concentrations. The procedural steps of cyanogen bromide cleavage,
CC oxidative sulphitolsis and related purification steps can thus be
CC eliminated, along with the use of high concentrations of mercaptan or the
CC use of hydrophobic absorbent resins
XX
SQ Sequence 52 AA;

Query Match 100.0%; Score 294; DB 2; Length 52;
Best Local Similarity 100.0%; Pred. No. 1.1e-26;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52

RESULT 2
AAR68901
ID AAR68901 standard; peptide; 56 AA.
XX

AC AAR68901;
XX
DT 25-MAR-2003 (revised)
DT 02-MAR-1995 (first entry)
XX
DE Human pro-insulin 3.
XX
KW Pro-insulin; A-chain; B-chain; C-chain; disulphide; mercaptan;
KW chaotropic agent.
XX
OS Homo sapiens.
XX
PN EP600372-A1.
XX
PD 08-JUN-1994.
XX
PF 25-NOV-1993; 93EP-00118993.
XX
PR 02-DEC-1992; 92DE-04240420.
XX
PA (FARH) HOECHST AG.
XX
PI Obermeier R, Gerl M, Ludwig J, Sabel W;
XX
DR WPI; 1994-177718/22.
XX
PT Prodn. of pro-insulin with correct di:sulphide bridges - by treating
PT recombinant precursor protein with mercaptan in alkali and in presence of
PT chaotropic agent, then isolation on hydrophobic resin.
XX
PS Disclosure; Page 12; 15pp; German.
XX
CC Pro-insulin is produced by treating recombinant precursor protein with a
CC mercaptan to provide 2-10 SH residues per Cys residue, in presence of a
CC chaotropic agent and in aq. medium of pH 10-11, treating the prod. with 3
CC -50 g hydrophobic adsorber resin per l aq. medium of pH 4-7, isolating
CC the adsorbed resin and pro-insulin and desorbing the pro-insulin. This
CC method produces pro-insulin with correctly bonded Cys bridges. Compared
CC with known methods it involves fewer stages (esp. no sulphitolsis or
CC cyanogen bromide cleavage) and overall losses during purification are
CC reduced, i.e. the process is quicker and gives better yields. Sequences
CC of insulin chain A, B and C are given in AAR68895-97. Sequences of pro-
CC insulin 1-4 are given in AAR68898-901. (Updated on 25-MAR-2003 to correct
CC PN field.)
XX
SQ Sequence 56 AA;

Query Match 100.0%; Score 294; DB 2; Length 56;
Best Local Similarity 100.0%; Pred. No. 1.1e-26;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
Db 5 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 56

RESULT 3

AAR78665

ID AAR78665 standard; protein; 56 AA.

XX

AC AAR78665;

XX

DT 03-APR-1996 (first entry)

XX

DE Proinsulin sequence 3.

XX

KW Proinsulin; post-translational modification; recombinant production;

KW protein folding; conformation.

XX

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Region 1. .4

FT /label= R2

FT /note= "a peptide of 4 amino acids"

FT Peptide 5. .34

FT /label= R1-(B2-B29)-Y

FT /note= "human insulin B-chain"

FT Region 35

FT /label= X

FT Peptide 36. .56

FT /label= Gly-(A2-A20)-R3

FT /note= "human insulin A-chain"

XX

PN EP668292-A2.

XX

PD 23-AUG-1995.

XX

PF 09-FEB-1995; 95EP-00101748.

XX

PR 18-FEB-1994; 94DE-04405179.

XX

PA (FARH) HOECHST AG.

XX

PI Obermeier R, Gerl M, Ludwig J, Sabel W;

XX

DR WPI; 1995-284754/38.

XX

PT Isolation of insulin that is correctly post-translationally processed -
PT by reacting pro:insulin with a mercaptan in the presence of a chaotropic
PT agent and purifcn. after absorption to hydrophobic resin.

XX

PS Example 2; Page 13; 16pp; German.

XX

CC The present sequence is an example of a proinsulin molecule corresp. to
CC the general formula R2-R1-(B2-B29)-Y-X-Gly-(A2-A20)-R3 (II). In formula
CC (II), X = Lys, Arg or a peptide of 2-35 amino acids contg. Lys or Arg at
CC the N- and C-termini; Y = a natural amino acid; R1 = Phe or a bond; R2 =
CC H, Arg, Lys, a peptide of 2-45 amino acids contg. Arg or Lys at the N-
CC and C-termini; R3 = a natural amino acid; (A2-A20) and (B2-B29) are the
CC insulin A- and B-chain sequences from human or other insulin. The
CC proinsulin molecule (produced in recombinant E.coli) is reacted with
CC mercaptan at a ratio of 2-10 SH residues of mercaptan per Cys residue of
CC proinsulin. The reaction takes place in the presence of a chaotropic

CC chaotropic agent and in aq. medium of pH 10-11, treating the prod. with 3
CC -50 g hydrophobic adsorber resin per l aq. medium of pH 4-7, isolating
CC the adsorbed resin and pro-insulin and desorbing the pro-insulin. This
CC method produces pro-insulin with correctly bonded Cys bridges. Compared
CC with known methods it involves fewer stages (esp. no sulphitolsis or
CC cyanogen bromide cleavage) and overall losses during purification are
CC reduced, i.e. the process is quicker and gives better yields. Sequences
CC of insulin chain A, B and C are given in AAR68895-97. Sequences of pro-
CC insulin 1-4 are given in AAR68898-901. (Updated on 25-MAR-2003 to correct
CC PN field.)

XX

SQ Sequence 96 AA;

Query Match 100.0%; Score 294; DB 2; Length 96;
Best Local Similarity 100.0%; Pred. No. 2e-26;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 45 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 96

RESULT 6

AAR78662

ID AAR78662 standard; protein; 96 AA.

XX

AC AAR78662;

XX

DT 03-APR-1996 (first entry)

XX

DE Fusion protein contg. proinsulin sequence 3.

XX

KW Proinsulin; post-translational modification; recombinant production;
KW protein folding; conformation.

XX

OS Synthetic.

XX

FH Key Location/Qualifiers
FT Region 41. .44
FT /label= R2
FT /note= "a peptide of 4 amino acids"
FT Peptide 45. .74
FT /label= R1-(B2-B29)-Y
FT /note= "human insulin B-chain"
FT Region 75
FT /label= X
FT Peptide 76. .96
FT /label= Gly-(A2-A20)-R3
FT /note= "human insulin A-chain"

XX

PN EP668292-A2.

XX

PD 23-AUG-1995.

XX

PF 09-FEB-1995; 95EP-00101748.

XX

PR 18-FEB-1994; 94DE-04405179.

XX
PA (FARH) HOECHST AG.
XX
PI Obermeier R, Gerl M, Ludwig J, Sabel W;
XX
DR WPI; 1995-284754/38.
XX
PT Isolation of insulin that is correctly post-translationally processed -
PT by reacting pro:insulin with a mercaptan in the presence of a chaotropic
PT agent and purificn. after absorption to hydrophobic resin.
XX
PS Example 2; Page 8; 16pp; German.
XX
CC The present sequence is that of a fusion protein, produced in E.coli
CC which contains an example of a proinsulin molecule corresp. to the
CC general formula R2-R1-(B2-B29)-Y-X-Gly-(A2-A20)-R3 (II). In formula (II),
CC X = Lys, Arg or a peptide of 2-35 amino acids contg. Lys or Arg at the N-
CC and C-termini; Y = a natural amino acid; R1 = Phe or a bond; R2 = H, Arg,
CC Lys, a peptide of 2-45 amino acids contg. Arg or Lys at the N- and C-
CC termini; R3 = a natural amino acid; (A2-A20) and (B2-B29) are the insulin
CC A- and B-chain sequences from human or other insulin. The proinsulin
CC molecule, released by cyanogen bromide, is reacted with mercaptan at a
CC ratio of 2-10 SH residues of mercaptan per Cys residue of proinsulin. The
CC reaction takes place in the presence of a chaotropic auxiliary agent at
CC pH 10-11 and results in proinsulin with correctly linked cystine bridges.
CC Reaction with trypsin and opt. carboxypeptidase B yields correctly folded
CC insulin. The insulin is isolated by absorption on a hydrophobic resin
XX
SQ Sequence 96 AA;

RESULT 7
AAY42860
ID AAY42860 standard; protein; 107 AA.
XX
AC AAY42860;
XX
DT 19-JAN-2000 (first entry)
XX
DE hGH-mini-proinsulin chimeric protein.
XX
KW Insulin; precursor; growth hormone; chaperone; intramolecular; folding;
KW conformation; chimeric protein; cleavable; recombinant; production;
KW yield.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9950302-A1.

XX
PD 07-OCT-1999.
XX
PF 31-MAR-1998; 98WO-CN000052.
XX
PR 31-MAR-1998; 98WO-CN000052.
XX
PA (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.
XX
PI Gan Z;
XX
DR WPI; 1999-610839/52.
XX
PT New chimeric proteins containing human growth hormone fragment, used
PT particularly for the production of human insulin.
XX
PS Claim 13; Page 30; 46pp; English.
XX
CC This sequence represents a chimeric protein, hGH-mini-proinsulin. This
CC chimeric protein contains an N-terminal fragment of human growth hormone
CC (hGH) of the sequence given in AAY42855, a cleavable peptide linker
CC (AAY42857), and a human insulin precursor comprising insulin A and B
CC chains (AAY42859). The hGH portion of the chimeric protein acts as an
CC intramolecular chaperone (IMC) for the insulin precursor, enabling it to
CC fold correctly. The cleavable peptide linker has a C-terminal Arg residue
CC which enables the hGH portion of the chimeric protein to be removed after
CC folding has taken place. Production of recombinant human insulin via an
CC hGH-proinsulin chimeric protein can provide human insulin with correctly
CC linked cysteine bridges with fewer necessary procedural steps, and hence
CC resulting in a higher yield of human insulin. The IMC sequences not only
CC protect insulin sequences from intracellular degradation by a
CC microorganism host, but also promote the folding of the fused insulin
CC precursor, facilitate the solubility of the fusion protein and decrease
CC the intermolecular interactions among the fusion proteins, thus allowing
CC folding of the fused insulin precursor at commercially useful high
CC concentrations. The procedural steps of cyanogen bromide cleavage,
CC oxidative sulphitolsis and related purification steps can thus be
CC eliminated, along with the use of high concentrations of mercaptan or the
CC use of hydrophobic absorbent resins
XX
SQ Sequence 107 AA;

Query Match 100.0%; Score 294; DB 2; Length 107;
Best Local Similarity 100.0%; Pred. No. 2.2e-26;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 56 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107

RESULT 8
AAR98897
ID AAR98897 standard; protein; 116 AA.
XX
AC AAR98897;
XX

DT 03-FEB-1997 (first entry)
XX
DE SOD-proinsulin hybrid polypeptide.
XX
KW Insulin; proinsulin; hybrid polypeptide; protein folding;
KW enzymatic cleavage; cyanogen bromide; sulphitolsis.
XX
OS Homo sapiens.
XX
PN WO9620724-A1.
XX
PD 11-JUL-1996.
XX
PF 29-DEC-1994; 94WO-US013268.
XX
PR 29-DEC-1994; 94WO-US013268.
XX
PA (BIOT-) BIO-TECHNOLOGY GENERAL CORP.
XX
PI Hartman JR, Mendelovitz S, Gorecki M;
XX
DR WPI; 1996-333766/33.
DR N-PSDB; AAT34670.
XX
PT Recombinant insulin prodn. by correctly folding pro-insulin hybrid
PT polypeptide - then enzymatic cleavage of folded product, does not require
PT sulphite protection of SH nor use of cyanogen bromide.
XX
PS Example 1B; Fig 7; 69pp; English.
XX
CC A new method for the production of recombinant human insulin comprises
CC folding a hybrid polypeptide comprising proinsulin under conditions that
CC permit correct disulphide bond formation and subjecting that folded
CC protein to enzymatic cleavage. The insulin produced can then be purified.
CC This sequence is a SOD-insulin B chain-Arg-insulin A chain hybrid
CC polypeptide and is encoded by the plasmid construct pDBAST-LAT.
CC Transformation of the proper E.coli host cells with pDBAST-LAT results in
CC the efficient expression of the proinsulin hybrid polypeptide, useful for
CC human insulin production. The method produces recombinant human insulin
CC identical to the natural hormone. Hazardous and cumbersome procedures
CC involving cyanogen bromide and sulphitolsis to protect SH groups are
CC avoided since the entire hybrid polypeptide folds efficiently to the
CC native structure even with the leader attached and Cys unprotected
XX
SQ Sequence 116 AA;

Query Match 100.0%; Score 294; DB 2; Length 116;
Best Local Similarity 100.0%; Pred. No. 2.4e-26;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFYTPTKTRGIVEQCCTSICSLYQLENYCN 52
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 65 FVNQHLCGSHLVEALYLVCGERGFYTPTKTRGIVEQCCTSICSLYQLENYCN 116

RESULT 9
AAR71692

ID AAR71692 standard; protein; 137 AA.
XX
AC AAR71692;
XX
DT 25-MAR-2003 (revised)
DT 20-NOV-1995 (first entry)
XX
DE Mating factor alpha 1-Insulin precursor ArgB31.
XX
KW Human insulin precursor ArgB31; diabetes; Zinc ion complex;
KW mating factor alpha 1.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Protein 1. .85
FT /label= mating factor alpha-1
FT Peptide 86. .116
FT /label= B-chain
FT Peptide 117. .137
FT /label= A-chain
XX
PN WO9507931-A1.
XX
PD 23-MAR-1995.
XX
PF 16-SEP-1994; 94WO-DK000347.
XX
PR 17-SEP-1993; 93DK-00001044.
PR 02-FEB-1994; 94US-00190829.
XX
PA (NOVO) NOVO-NORDISK AS.
XX
PI Havelund S, Halstrom JB, Jonassen I, Andersen AS, Markussen J;
XX
DR WPI; 1995-131314/17.
DR N-PSDB; AAQ86425.
XX
PT Acylated insulin deriv. which may be present as a Zinc ion complex - is
PT used to treat diabetes and is rapid acting.
XX
PS Example 5; Page 78; 100pp; English.
XX
CC AAQ86425 encodes AAR71692 mating factor alpha 1-Insulin precursor ArgB31.
CC ArgB31 comprises the B and A chains of a claimed human insulin
CC derivative. In the final claimed compsn. they are covalently connected
CC via disulphide bonds between Cys residues A7/B7 and A20/B19. The
CC derivative, which may be present as a zinc ion complex, can be used as a
CC fast action treatment for diabetes. (Updated on 25-MAR-2003 to correct PN
CC field.)
XX
SQ Sequence 137 AA;

Query Match 100.0%; Score 294; DB 2; Length 137;
Best Local Similarity 100.0%; Pred. No. 2.8e-26;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 86 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 137

RESULT 10

AAR71694

ID AAR71694 standard; protein; 145 AA.

XX

AC AAR71694;

XX

DT 25-MAR-2003 (revised)

DT 20-NOV-1995 (first entry)

XX

DE Mating factor alpha 1-Insulin precursor ArgB1, ArgB31 N-terminal.

XX

KW Human insulin precursor ArgB1, ArgB31; diabetes; Zinc ion complex;
KW mating factor alpha 1; N-terminal EEAEAEAR.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Protein 1. .85

FT /label= mating factor alpha-1

FT Peptide 86. .93

FT /label= N-terminal peptide

FT Peptide 94. .124

FT /label= B-chain

FT Peptide 125. .145

FT /label= A-chain

XX

PN WO9507931-A1.

XX

PD 23-MAR-1995.

XX

PF 16-SEP-1994; 94WO-DK000347.

XX

PR 17-SEP-1993; 93DK-00001044.

PR 02-FEB-1994; 94US-00190829.

XX

PA (NOVO) NOVO-NORDISK AS.

XX

PI Havelund S, Halstrom JB, Jonassen I, Andersen AS, Markussen J;

XX

DR WPI; 1995-131314/17.

DR N-PSDB; AAQ86429.

XX

PT Acylated insulin deriv. which may be present as a Zinc ion complex - is
PT used to treat diabetes and is rapid acting.

XX

PS Example 5; Page 82-83; 100pp; English.

XX

CC AAQ86429 encodes AAR71694 mating factor alpha 1-Insulin precursor ArgB1,
CC ArgB31 N-terminal EEAEAEAR. The insulin precursor comprises the B and A
CC chains of a claimed human insulin derivative preceded by the N-terminal
CC amino acids EEAEAEAR. In the final claimed compsn. they are covalently
CC connected via disulphide bonds between Cys residues A7/B7 and A20/B19.

PT Acylated insulin deriv. which may be present as a Zinc ion complex - is used to treat diabetes and is rapid acting.
XX
PS Example 6; Page 85; 100pp; English.
XX
CC AAQ86432 encodes AAR71695 mating factor alpha 1-Insulin precursor ArgB1,
CC ArgB31 N-terminal EEAEEAER. The insulin precursor comprises the B and A
CC chains of a claimed human insulin derivative preceded by the N-terminal
CC amino acids EEAEEAER. In the final claimed compsn. they are covalently
CC connected via disulphide bonds between Cys residues A7/B7 and A20/B19.
CC The derivative, which may be present as a zinc ion complex, can be used
CC as a fast action treatment for diabetes. (Updated on 25-MAR-2003 to
CC correct PN field.)
XX
SQ Sequence 146 AA;

Query Match 100.0%; Score 294; DB 2; Length 146;
Best Local Similarity 100.0%; Pred. No. 3e-26;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 95 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 146

RESULT 12
AAY42861
ID AAY42861 standard; protein; 150 AA.
XX
AC AAY42861;
XX
DT 19-JAN-2000 (first entry)
XX
DE Chimeric protein, SEQ ID 7.
XX
KW Insulin; precursor; growth hormone; chaperone; intramolecular; folding;
KW conformation; chimeric protein; cleavable; recombinant; production;
KW yield.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9950302-A1.
XX
PD 07-OCT-1999.
XX
PF 31-MAR-1998; 98WO-CN000052.
XX
PR 31-MAR-1998; 98WO-CN000052.
XX
PA (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.
XX
PI Gan Z;
XX
DR WPI; 1999-610839/52.
XX
PT New chimeric proteins containing human growth hormone fragment, used

PT particularly for the production of human insulin.

XX *Journal of English Linguistics*

PS

XX
CC This sequence represents a chimeric protein, which contains an N-terminal
CC fragment of human growth hormone (hGH) of the sequence given in AAY42856,
CC a cleavable peptide linker (AAV42857), and a human insulin precursor
CC comprising insulin A and B chains (AAV42859). The hGH portion of the
CC chimeric protein acts as an intramolecular chaperone (IMC) for the
CC insulin precursor, enabling it to fold correctly. The cleavable peptide
CC linker has a C-terminal Arg residue which enables the hGH portion of the
CC chimeric protein to be removed after folding has taken place. Production
CC of recombinant human insulin via an hGH-proinsulin chimeric protein can
CC provide human insulin with correctly linked cysteine bridges with fewer
CC necessary procedural steps, and hence resulting in a higher yield of
CC human insulin. The IMC sequences not only protect insulin sequences from
CC intracellular degradation by a microorganism host, but also promote the
CC folding of the fused insulin precursor, facilitate the solubility of the
CC fusion protein and decrease the intermolecular interactions among the
CC fusion proteins, thus allowing folding of the fused insulin precursor at
CC commercially useful high concentrations. The procedural steps of cyanogen
CC bromide cleavage, oxidative sulphitolsis and related purification steps
CC can thus be eliminated, along with the use of high concentrations of
CC mercaptan or the use of hydrophobic absorbent resins

xx

SQ Sequence 150 AA;

RESULT 13

AAR04582

ID AAR04582 standard; protein; 57 AA.

xx

AC AAR04582;

xx

DT 25-MAR-2003 (revised)

DT 14-SEP-1990 (first entry)

xx

DE Proinsulin analogue with a Lys residue linking the A and B chains.

xx

KW insulin fusion protein, pro-insulin B-chain, ds.

RW Wys Wys bridge, ab.
XX

05

xx

FK

FT Peptide 1. .35

FT /label

FT Misc-difference 36

FT / 1.

FT Peptide 37. .57
FT /label= Insulin A chain
XX
PN EP367163-A.
XX
PD 09-MAY-1990.
XX
PF 28-OCT-1989; 89EP-00120056.
XX
PR 03-NOV-1988; 88DE-03837273.
PR 19-AUG-1989; 89DE-03927449.
XX
PA (FARH) HOECHST AG.
XX
PI Koller KP, Riess GJ, Uhlmann E, Wallmeier H;
XX
DR WPI; 1990-141149/19.
DR N-PSDB; AAQ04335.
XX
PT New insulin fusion proteins - comprise pro-insulin analogue linked to
PT tendamistate.
XX
PS Disclosure; Page ?; -pp; German.
XX
CC This sequence is joined to the C-terminus of an N-terminal fragment
CC comprising opt. modified tendamistate. This fusion protein may be
CC converted into human insulin using known methods. The synthetic gene was
CC prepared by the phosphoramidite method. See also AAQ04336. (Updated on 25
CC -MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to correct PI
CC field.)
XX
SQ Sequence 57 AA;

Query Match 99.0%; Score 291; DB 2; Length 57;
Best Local Similarity 98.1%; Pred. No. 2.6e-26;
Matches 51; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 6 FVNQHLCGSHLVEALYLVCGERGFFYTPKTKGIVEQCCTSICSLYQLENYCN 57

RESULT 14
AAR11899
ID AAR11899 standard; protein; 52 AA.
XX
AC AAR11899;
XX
DT 25-MAR-2003 (revised)
DT 22-JUL-1991 (first entry)
XX
DE Example of human insulin precursor.
XX
KW Human insulin; diabetes; transpeptidation.
XX
OS Homo sapiens.
XX

PN EP427296-A.
XX
PD 15-MAY-1991.
XX
PF 29-MAY-1985; 90EP-00121887.
XX
PR 30-MAY-1984; 84DK-00002665.
PR 08-FEB-1985; 85DK-00000582.
XX
PA (NOVO) NOVO-NORDISK AS.
XX
PI Markussen J, Fiil N, Ammerer G, Hansen MT, Thim L, Norris K;
PI Voigt HO;
XX
DR WPI; 1991-141828/20.
XX
PT Human insulin precursors - expressed with correctly positioned
PT di:sulphide bridges giving improved resistance to proteolysis.
XX
PS Claim 3; Page 18; 28pp; English.
XX
CC This human insulin precursor has correctly positioned disulphide bridges
CC between the A and B chains and is more resistant to proteolytic digestion
CC than prior art insulin precursors. Yeast strains transformed with DNA
CC encoding this precursor can be cultured to secrete it in high yields. The
CC precursor can be converted into mature human insulin by transpeptidation.
CC See also AAR11897-98. (Updated on 25-MAR-2003 to correct PF field.)
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 52 AA;

```

Query Match          97.6%;  Score 287;  DB 2;  Length 52;
Best Local Similarity  96.2%;  Pred. No. 6.8e-26;
Matches   50;  Conservative   2;  Mismatches   0;  Indels   0;  Gaps   0;

Y      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYNCN 52
       |||||||:|||||||:|||||||:|||||||:|||||||:|||||||:|||||||:|||||||
b      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKSKKGIVEQCCTSICSLYQLENYNCN 52

```

RESULT 15
AAR65883
ID AAR65883 standard; protein; 53 AA.
XX
AC AAR65883;
XX
DT 16-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)
XX
DE Di-Arg-(B31-32)-Human insulin amorphous, monospherical deriv.
XX
KW Human insulin; recombinant production; amorphous; monospherical form;
KW diabetes mellitus.
XX
OS Homo sapiens; (produced recombinantly in Escherichia coli).
XX

FH Key Location/Qualifiers
 FT Protein 1. .30
 FT /label= insulin_B-chain
 FT Protein 33. .53
 FT /label= insulin_A-chain
 XX
 PN EP622376-A1.
 XX
 PD 02-NOV-1994.
 XX
 PF 21-APR-1994; 94EP-00106196.
 XX
 PR 27-APR-1993; 93DE-04313702.
 XX
 PA (FARH) HOECHST AG.
 XX
 PI Obermeier R, Sabel W, Deil P, Geisen K;
 XX
 DR WPI; 1994-334579/42.
 XX
 PT Amorphous, mono-spherical form of insulin derivs. - for treating diabetes
 PT mellitus, are produced by diluting soln. in aq. isopropanol, are stable
 PT when dried or in suspension.
 XX
 PS Example 2; Page 5; 10pp; German.
 XX
 CC This sequence is a specific example of an insulin derivative which can be
 CC obtained in amorphous, monospherical form by dissolving in an n-
 CC propanol/buffer mixture (pH 4.5-6.5) having n-propanol content 15%
 CC relative to water. The solution is then diluted with water to reduce n-
 CC propanol content to below 15%. The resulting insulin preparation is
 CC stable and can be used for the treatment of diabetes mellitus. (Updated
 CC on 25-MAR-2003 to correct PN field.) (Updated on 16-OCT-2003 to
 CC standardise OS field)
 XX
 SQ Sequence 53 AA;

 Query Match 96.4%; Score 283.5; DB 2; Length 53;
 Best Local Similarity 98.1%; Pred. No. 1.8e-25;
 Matches 52; Conservative 0; Mismatches 0; Indels 1; Gaps 1

 Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT-RGIVEQCCTSICSLYQLENYCN 52
 Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRRGIVEQCCTSICSLYQLENYCN 53

Search completed: July 15, 2004, 16:35:34
Job time : 28.1642 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 16:30:45 ; Search time 7.85821 Seconds
(without alignments)
341.624 Million cell updates/sec

Title: US-09-423-100-5
Perfect score: 294
Sequence: 1 FVNQHLCGSHLVEALYLVCG.....IVEQCCTSICSLYQLENYCN 52

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query				Description
		Match	Length	DB	ID	
1	294	100.0	56	1	US-08-160-376A-7	Sequence 7, Appli
2	294	100.0	56	1	US-08-389-487-11	Sequence 11, Appli
3	294	100.0	63	1	US-08-160-376A-6	Sequence 6, Appli
4	294	100.0	66	1	US-08-291-060B-5	Sequence 5, Appli
5	294	100.0	96	1	US-08-160-376A-5	Sequence 5, Appli
6	294	100.0	96	1	US-08-389-487-8	Sequence 8, Appli
7	294	100.0	137	1	US-08-400-256-39	Sequence 39, Appli
8	294	100.0	137	3	US-08-975-365-39	Sequence 39, Appli
9	294	100.0	145	1	US-08-400-256-45	Sequence 45, Appli
10	294	100.0	145	3	US-08-975-365-45	Sequence 45, Appli
11	294	100.0	146	1	US-08-400-256-48	Sequence 48, Appli

12	294	100.0	146	3	US-08-975-365-48	Sequence 48, Appl
13	291	99.0	57	1	US-08-030-731A-44	Sequence 44, Appl
14	283.5	96.4	53	1	US-08-233-617-4	Sequence 4, Appl
15	283.5	96.4	53	4	US-08-981-988A-42	Sequence 42, Appl
16	278.5	94.7	51	4	US-09-477-924-3	Sequence 3, Appl
17	278.5	94.7	51	4	US-09-723-981-3	Sequence 3, Appl
18	278.5	94.7	51	4	US-09-723-896-3	Sequence 3, Appl
19	277.5	94.4	53	1	US-08-233-617-3	Sequence 3, Appl
20	277	94.2	65	3	US-08-900-574-3	Sequence 3, Appl
21	276.5	94.0	55	3	US-08-900-574-6	Sequence 6, Appl
22	276.5	94.0	66	3	US-08-900-574-5	Sequence 5, Appl
23	276.5	94.0	67	4	US-08-981-988A-1	Sequence 1, Appl
24	276.5	94.0	67	4	US-08-981-988A-5	Sequence 5, Appl
25	276	93.9	67	3	US-08-900-574-7	Sequence 7, Appl
26	275.5	93.7	53	3	US-09-261-853-2	Sequence 2, Appl
27	275.5	93.7	65	1	US-08-468-674B-71	Sequence 71, Appl
28	275.5	93.7	65	1	US-08-780-571-71	Sequence 71, Appl
29	275.5	93.7	89	1	US-08-468-674B-41	Sequence 41, Appl
30	275.5	93.7	89	1	US-08-780-571-41	Sequence 41, Appl
31	275.5	93.7	91	1	US-08-468-674B-45	Sequence 45, Appl
32	275.5	93.7	91	1	US-08-780-571-45	Sequence 45, Appl
33	275.5	93.7	104	1	US-08-400-256-15	Sequence 15, Appl
34	275.5	93.7	104	3	US-08-975-365-15	Sequence 15, Appl
35	275.5	93.7	117	3	US-09-012-669F-37	Sequence 37, Appl
36	275.5	93.7	124	1	US-08-446-646-3	Sequence 3, Appl
37	275.5	93.7	124	3	US-09-012-669F-36	Sequence 36, Appl
38	275.5	93.7	138	3	US-08-932-082-19	Sequence 19, Appl
39	275.5	93.7	138	4	US-09-861-687-19	Sequence 19, Appl
40	275.5	93.7	140	1	US-08-400-256-33	Sequence 33, Appl
41	275.5	93.7	140	1	US-08-400-256-42	Sequence 42, Appl
42	275.5	93.7	140	3	US-08-975-365-33	Sequence 33, Appl
43	275.5	93.7	140	3	US-08-975-365-42	Sequence 42, Appl
44	273.5	93.0	67	4	US-08-981-988A-2	Sequence 2, Appl
45	272.5	92.7	53	3	US-08-900-574-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
 US-08-160-376A-7
 ; Sequence 7, Application US/08160376A
 ; Patent No. 5473049
 ; GENERAL INFORMATION:
 ; APPLICANT: Obermeier, Ranier
 ; APPLICANT: Gerl, Martin
 ; APPLICANT: Ludwig, Jurgen
 ; APPLICANT: Sabel, Walter
 ; TITLE OF INVENTION: Process For Obtaining Proinsulin
 ; TITLE OF INVENTION: Possessing Correctly Linked
 ; TITLE OF INVENTION: Cystine Bridges
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Kenneth A. Genoni, Esq.
 ; STREET: Rt. 202-206 No. 5473049th/P.O. Box 2500
 ; CITY: Somerville
 ; STATE: New Jersey

; COUNTRY: U.S.A.
; ZIP: 08876-1258
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: IBM 386
; OPERATING SYSTEM: WINDOWS 3.1
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/160,376A
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GE P 4240420.7
; FILING DATE: December 2, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Barbara V. Maurer, Esq.
; REGISTRATION NUMBER: 31,287
; REFERENCE/DOCKET NUMBER: HOE 92/F 384
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 231-4079
; TELEFAX: (908) 231-2255
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56 Amino Acids
; TYPE: Amino Acid (AA)
; TOPOLOGY: not relevant

US-08-160-376A-7

Query Match 100.0%; Score 294; DB 1; Length 56;
Best Local Similarity 100.0%; Pred. No. 7.3e-29;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
Db 5 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 56

RESULT 2
US-08-389-487-11
; Sequence 11, Application US/08389487
; Patent No. 5663291
; GENERAL INFORMATION:
; APPLICANT: Obermeier, Rainer
; APPLICANT: Gerl, Martin
; APPLICANT: Ludwig, Jurgen
; APPLICANT: Sabel, Walter
; TITLE OF INVENTION: Process for Obtaining Insulin Having
; TITLE OF INVENTION: Correctly Linked Cystine Bridges
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20005-3315

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/389,487
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Einaudi, Carol P.
; REGISTRATION NUMBER: 32,220
; REFERENCE/DOCKET NUMBER: 02481.1424-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-389-487-11

Query Match 100.0%; Score 294; DB 1; Length 56;
Best Local Similarity 100.0%; Pred. No. 7.3e-29;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 56

RESULT 3
US-08-160-376A-6
; Sequence 6, Application US/08160376A
; Patent No. 5473049
; GENERAL INFORMATION:
; APPLICANT: Obermeier, Ranier
; APPLICANT: Gerl, Martin
; APPLICANT: Ludwig, Jurgen
; APPLICANT: Sabel, Walter
; TITLE OF INVENTION: Process For Obtaining Proinsulin
; TITLE OF INVENTION: Possessing Correctly Linked
; TITLE OF INVENTION: Cystine Bridges
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth A. Genoni, Esq.
; STREET: Rt. 202-206 No. 5473049th/P.O. Box 2500
; CITY: Somerville
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 08876-1258
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: IBM 386

OPERATING SYSTEM: WINDOWS 3.1
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/160,376A
FILING DATE: December 1, 1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GE P 4240420.7
FILING DATE: December 2, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Barbara V. Maurer, Esq.
REGISTRATION NUMBER: 31,287
REFERENCE/DOCKET NUMBER: HOE 92/F 384
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 231-4079
TELEFAX: (908) 231-2255
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 63 Amino Acids
TYPE: Amino Acid (AA)
TOPOLOGY: not relevant

RESULT 4

US-08-291-060B-5

; Sequence 5, Application US/08291060B

; Patent No. 5728543

; GENERAL INFORMATION:

; APPLICANT: Dorschug, Michael

; APPLICANT: Koller, Klaus-Peter

; APPLICANT: Marquardt, Rudiger

; APPLICANT: Meiwes, Johannes

; TITLE OF INVENTION: An Enzymatic Process for the

; TITLE OF INVENTION: Cor

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Hender

; ADDRESSEE: Dunne

; STREET: 130

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005-3315

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; SYSTEM: PC DOS/MS DOS

; OPERATING SYSTEM: PC-DOS/MS-DOS
; AUTOMATION: Full Auto Backup Version #1.0 Version #1.30

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/291,060B
; FILING DATE: 08-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Einaudi, Carol P.
; REGISTRATION NUMBER: 32,220
; REFERENCE/DOCKET NUMBER: 02481.1105-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4366
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 66 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

US-08-291-060B-5

Query Match 100.0%; Score 294; DB 1; Length 66;
Best Local Similarity 100.0%; Pred. No. 8.8e-29;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYC 52
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYC 66

RESULT 5
US-08-160-376A-5
; Sequence 5, Application US/08160376A
; Patent No. 5473049
; GENERAL INFORMATION:
; APPLICANT: Obermeier, Ranier
; APPLICANT: Gerl, Martin
; APPLICANT: Ludwig, Jurgen
; APPLICANT: Sabel, Walter
; TITLE OF INVENTION: Process For Obtaining Proinsulin
; TITLE OF INVENTION: Possessing Correctly Linked
; TITLE OF INVENTION: Cystine Bridges
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth A. Genoni, Esq.
; STREET: Rt. 202-206 No. 5473049th/P.O. Box 2500
; CITY: Somerville
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 08876-1258
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: IBM 386
; OPERATING SYSTEM: WINDOWS 3.1
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/160,376A
; FILING DATE: December 1, 1993

; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GE P 4240420.7
; FILING DATE: December 2, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Barbara V. Maurer, Esq.
; REGISTRATION NUMBER: 31,287
; REFERENCE/DOCKET NUMBER: HOE 92/F 384
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 231-4079
; TELEFAX: (908) 231-2255
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 96 Amino Acids
; TYPE: Amino Acid (AA)
; TOPOLOGY: not relevant

US-08-160-376A-5

Query Match 100.0%; Score 294; DB 1; Length 96;
Best Local Similarity 100.0%; Pred. No. 1.3e-28;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
Db 45 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 96

RESULT 6

US-08-389-487-8

; Sequence 8, Application US/08389487
; Patent No. 5663291
; GENERAL INFORMATION:
; APPLICANT: Obermeier, Rainer
; APPLICANT: Gerl, Martin
; APPLICANT: Ludwig, Jurgen
; APPLICANT: Sabel, Walter
; TITLE OF INVENTION: Process for Obtaining Insulin Having
; TITLE OF INVENTION: Correctly Linked Cystine Bridges
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/389,487
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:

; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 137 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-975-365-39

RESULT 9
US-08-400-256-45
; Sequence 45, Application US/08400256
; Patent No. 5750497
; GENERAL INFORMATION:
; APPLICANT: Havelund, Svend
; APPLICANT: Halstrom, John
; APPLICANT: Jonassen, Ib
; APPLICANT: Andersen, Asser Sloth
; APPLICANT: Markussen, Jan
; TITLE OF INVENTION: ACYLATED INSULIN
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5750497o No. 5750497disk of No. 5750497th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/400,256
; FILING DATE: 03-MAR-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3985.220-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 145 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
| |||||||
Db 95 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 146

RESULT 12
US-08-975-365-48
; Sequence 48, Application US/08975365
; Patent No. 6011007
; GENERAL INFORMATION:
; APPLICANT: Havelund, Svend
; APPLICANT: Halstrom, John
; APPLICANT: Jonassen, Ib
; APPLICANT: Andersen, Asser Sloth
; APPLICANT: Markussen, Jan
; TITLE OF INVENTION: ACYLATED INSULIN
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6011007o No. 6011007disk of No. 6011007th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/975,365
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/400,256
; FILING DATE: 03-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3985.220-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 146 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-975-365-48
Query Match 100.0%; Score 294; DB 3; Length 146;
Best Local Similarity 100.0%; Pred. No. 2e-28;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
| |||||||

RESULT 13
US-08-030-731A-44
; Sequence 44, Application US/08030731A
; Patent No. 5426036
; GENERAL INFORMATION:
; APPLICANT: Koller, Klaus-Peter
; APPLICANT: Riess, Guenther Johannes
; APPLICANT: Uhlmann, Eugen
; APPLICANT: Wallmeier, Holger
; TITLE OF INVENTION: Processes for the Preparation of Foreign
; TITLE OF INVENTION: Proteins in Streptomyces
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/030,731A
; FILING DATE: 12-MAR-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/189,840
; FILING DATE: 03-MAY-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/430,622
; FILING DATE: 01-NOV-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/687,610
; FILING DATE: 19-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/735,757
; FILING DATE: 29-JUL-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 37 14 866.4
; FILING DATE: 05-MAY-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 38 37 273.8
; FILING DATE: 03-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 39 27 449.7
; FILING DATE: 19-AUG-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 40 12 818.0
; FILING DATE: 21-APR-1990
; ATTORNEY/AGENT INFORMATION:

; NAME: Kirschner Michael K.
; REGISTRATION NUMBER: 34,851
; REFERENCE/DOCKET NUMBER: 02481-0593-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 57 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-030-731A-44

Query Match 99.0%; Score 291; DB 1; Length 57;
Best Local Similarity 98.1%; Pred. No. 1.7e-28;
Matches 51; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6 FVNQHLCGSHLVEALYLVCGERGFFYTPKTKGIVEQCCTSICSLYQLENYCN 57

RESULT 14
US-08-233-617-4
; Sequence 4, Application US/08233617
; Patent No. 5466666
; GENERAL INFORMATION:
; APPLICANT: Obermeier, Rainer
; APPLICANT: Sabel, Walter
; APPLICANT: Deil, Peter
; APPLICANT: Geisen, Karl
; TITLE OF INVENTION: Amorphous Monospherical Forms of Insulin
; TITLE OF INVENTION: Derivatives
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,617
; FILING DATE: 25-APR-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P 43 13 702.4
; FILING DATE: 27-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Carol P. Einaudi

US-08-981-988A-42

Query Match 96.4%; Score 283.5; DB 4; Length 53;
Best Local Similarity 98.1%; Pred. No. 1.3e-27;
Matches 52; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT-RGIVEQCCTSICSLYQLENYCN 52
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTKRGIVEQCCTSICSLYQLENYCN 53

Search completed: July 15, 2004, 16:42:32

Job time : 8.85821 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 16:29:19 ; Search time 5.8209 Seconds
(without alignments)
859.311 Million cell updates/sec

Title: US-09-423-100-5

Perfect score: 294

Sequence: 1 FVNQHLCGSHLVEALYLVCG.....IVEQCCTSICSLYQLENYCN 52

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			Description
		Match	Length	DB ID	
1	273.5	93.0	51	1 INWHP	insulin - sperm wh
2	273.5	93.0	51	1 INWHF	insulin - finback
3	273.5	93.0	51	1 INEL	insulin - elephant
4	273	92.9	96	2 PC7082	epidermal growth f
5	271.5	92.3	51	1 INHY	insulin - hamster
6	268.5	91.3	51	1 INMSSP	insulin - Egyptian
7	267.5	91.0	51	2 A59151	insulin precursor
8	267	90.8	110	1 IPHU	insulin precursor
9	267	90.8	110	2 B42179	insulin precursor
10	267	90.8	110	2 A42179	insulin precursor
11	267	90.8	110	2 JQ0178	insulin precursor
12	263.5	89.6	51	1 INWH1S	insulin - sei whal
13	263.5	89.6	51	1 INGT	insulin - goat

14	263.5	89.6	51	1	INCMA	insulin - Arabian
15	263	89.5	84	1	IPPG	insulin precursor
16	263	89.5	110	1	INRB	insulin precursor
17	262.5	89.3	51	1	INCT	insulin - cat
18	262	89.1	110	1	IPDG	insulin precursor
19	261.5	88.9	51	1	INMKSQ	insulin - common s
20	260	88.4	110	2	I48166	insulin precursor
21	258.5	87.9	105	1	IPBO	insulin precursor
22	256.5	87.2	51	2	JQ0362	insulin - North Am
23	252.5	85.9	77	1	INSH	insulin precursor
24	252	85.7	86	1	IPHO	insulin precursor
25	251.5	85.5	51	1	INCB	insulin - Chinchil
26	251	85.4	108	2	A39883	insulin precursor
27	250	85.0	108	1	INMS1	insulin 1 precurso
28	249	84.7	110	1	IPRT1	insulin 1 precurso
29	248.5	84.5	51	1	INGS	insulin - goose
30	248	84.4	110	1	IPRT2	insulin 2 precurso
31	248	84.4	110	1	INMS2	insulin 2 precurso
32	246	83.7	52	2	S44469	insulin I1 - North
33	246	83.7	52	2	S44470	insulin I2 - North
34	244.5	83.2	51	1	INPQ	insulin - crested
35	244.5	83.2	51	1	INTK	insulin - turkey (
36	244.5	83.2	51	1	INOS	insulin - ostrich
37	244.5	83.2	51	1	A61129	insulin - black-be
38	244.5	83.2	51	2	A60414	insulin - slider t
39	238.5	81.1	103	2	I51221	insulin precursor
40	238	81.0	52	2	S61361	insulin - Amphiuma
41	235.5	80.1	51	2	S63590	insulin - duckbill
42	234.5	79.8	107	1	IPCH	insulin precursor
43	233.5	79.4	81	1	IPDK	insulin precursor
44	231.5	78.7	51	1	INAQ	insulin - American
45	231	78.6	52	1	INGXA	insulin - alligato

ALIGNMENTS

RESULT 1
 INWHP
 insulin - sperm whale
 C;Species: Physeter catodon (sperm whale)
 C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
 C;Accession: A93142; A90082
 R;Ishihara, Y.; Saito, T.; Ito, Y.; Fujino, M.
Nature 181, 1468-1469, 1958
 A;Title: Structure of sperm- and sei-whale insulins and their breakdown by whale
 pepsin.
 A;Reference number: A93142
 A;Accession: A93142
 A;Molecule type: protein
 A;Residues: 1-30;31-51 <ISH>
 R;Harris, J.I.; Sanger, F.; Naughton, M.A.
Arch. Biochem. Biophys. 65, 427-428, 1956
 A;Title: Species differences in insulin.
 A;Reference number: A90082
 A;Accession: A90082
 A;Molecule type: protein

A;Residues: 1-30;31-51 <HAR>
C;Superfamily: insulin
C;Keywords: hormone; pancreas
F;1-30/Domain: insulin chain B #status experimental <BCN>
F;1-30,31-51/Product: insulin #status experimental <MAT>
F;31-51/Domain: insulin chain A #status experimental <ACH>
F;7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 93.0%; Score 273.5; DB 1; Length 51;
Best Local Similarity 96.2%; Pred. No. 1.5e-24;
Matches 50; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCTSICSLYQLENYCN 51

RESULT 2
INWHF
insulin - finback whale (tentative sequence)
C;Species: Balaenoptera physalus (finback whale, common rorqual)
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C;Accession: A91918
R;Hama, H.; Titani, K.; Sakaki, S.; Narita, K.
J. Biochem. 56, 285-293, 1964
A;Title: The amino acid sequence in fin-whale insulin.
A;Reference number: A91918
A;Accession: A91918
A;Molecule type: protein
A;Residues: 1-30;31-51 <HAM>
C;Superfamily: insulin
C;Keywords: hormone; pancreas
F;1-30/Domain: insulin chain B #status experimental <BCN>
F;1-30,31-51/Product: insulin #status experimental <MAT>
F;31-51/Domain: insulin chain A #status experimental <ACH>
F;7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 93.0%; Score 273.5; DB 1; Length 51;
Best Local Similarity 96.2%; Pred. No. 1.5e-24;
Matches 50; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCTSICSLYQLENYCN 51

RESULT 3
INEL
insulin - elephant
C;Species: Elephantidae gen. sp. (elephant)
C;Date: 24-Apr-1984 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999
C;Accession: A01584
R;Smith, L.F.
Am. J. Med. 40, 662-666, 1966
A;Title: Species variation in the amino acid sequence of insulin.
A;Reference number: A90029; MUID:66160119; PMID:5949593
A;Accession: A01584

A;Molecule type: protein
A;Residues: 1-30;31-51 <SMI>
A;Note: the species of elephant is not given, but it is most probably the Indian elephant (*Elephas maximus*)
C;Superfamily: insulin
C;Keywords: hormone; pancreas
F;1-30/Domain: insulin chain B #status experimental <BCN>
F;1-30,31-51/Product: insulin #status experimental <MAT>
F;31-51/Domain: insulin chain A #status experimental <ACH>
F;7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 93.0%; Score 273.5; DB 1; Length 51;
Best Local Similarity 94.2%; Pred. No. 1.5e-24;
Matches 49; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFYTPKTRGIVEQCCTSICSLYQLENYCN 52
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 FVNQHLCGSHLVEALYLVCGERGFYTPKT-GIVEQCCTGVCSLYQLENYCN 51

RESULT 4

PC7082
epidermal growth factor/single chain insulin fusion protein - *Bacillus brevis*
(fragment)

C; Species: *Bacillus brevis*

C:\Date: 18-Aug-2000 #sequence revision 18-Aug-2000 #text change 31-Mar-2003

C;Accession: PC7082; PC7083

R; Koh, M.; Hanaqata, H.; Ebisu, S.; Morihara, K.; Takagi, H.

Biosci. Biotechnol. Biochem. 64, 1079-1081, 2000

A;Title: Use of *Bacillus brevis* for synthesis and secretion of Des-B30 single-chain human insulin precursor.

A;Reference number: PC7082; MUID:20335834; PMID:10879487

A; Accession: PC7082

A;Molecule type: DNA

A: Residues: 1-96 <KOH>

A:Accession: PC7083

A: Molecule type: protein

A:Residues: 19-28 <KO2>

C: Genetics:

A:Gene: egf-sci

C:Superfamily: insulin

Query Match 92.9%; Score 273; DB 2; Length 96;

Best Local Similarity 96.2%: Pred. No. 2.9e-24;

Matches 50; Conservative 0; Mismatches 0; Indels 2; Gaps 1

QY 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 47 FVNQHLCGSHLVEALYLVCGERGFYTPK--GIVEQCCTSICSLYQLENYCN 96

RESULT 5

INHY

insulin - hamster

C; Species: Cricetinae gen. sp. (hamster)

C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999

C;Accession: A91456

R;Neelon, F.A.; Delcher, H.K.; Steinman, H.; Lebovitz, H.E.
Fed. Proc. 32, 300, 1973
A;Title: Structure of hamster insulin: comparison with a tumor insulin.
A;Reference number: A91456
A;Accession: A91456
A;Molecule type: protein
A;Residues: 1-30;31-51 <NEE>
C;Superfamily: insulin
C;Keywords: hormone; pancreas
F;1-30/Domain: insulin chain B #status experimental <BCN>
F;1-30,31-51/Product: insulin #status experimental <MAT>
F;31-51/Domain: insulin chain A #status experimental <ACH>
F;7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 92.3%; Score 271.5; DB 1; Length 51;
Best Local Similarity 94.2%; Pred. No. 2.5e-24;
Matches 49; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKS-GIVDQCCTSICSLYQLENYCN 51

RESULT 6
INMSSP
insulin - Egyptian spiny mouse (tentative sequence)
C;Species: Acomys cahirinus (Egyptian spiny mouse)
C;Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 31-Mar-2000
C;Accession: A01591
R;Buenzli, H.F.; Humbel, R.E.
Hoppe-Seyler's Z. Physiol. Chem. 353, 444-450, 1972
A;Title: Isolation and partial structural analysis of insulin from mouse (Mus musculus) and spiny mouse (Acomys cahirinus).
A;Reference number: A01591; MUID:72189454; PMID:5028210
A;Contents: composition
A;Accession: A01591
A;Molecule type: protein
A;Residues: 1-30;31-51 <BU>
C;Superfamily: insulin
C;Keywords: hormone; pancreas
F;1-30/Domain: insulin chain B #status predicted <BCN>
F;1-30,31-51/Product: insulin #status predicted <MAT>
F;31-51/Domain: insulin chain A #status predicted <ACH>
F;7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 91.3%; Score 268.5; DB 1; Length 51;
Best Local Similarity 92.3%; Pred. No. 5.5e-24;
Matches 48; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1 FVBQHLCGSHLVEALYLVCGERGFFYTPKS-GIVDQCCTSICSLYQLENYCN 51

RESULT 7
A59151
insulin precursor - jack bean (fragments)

N;Alternate names: hypoglycemic agent; plant insulin
C;Species: Canavalia ensiformis (jack bean)
C;Date: 07-Dec-1999 #sequence_revision 07-Dec-1999 #text_change 10-Dec-1999
C;Accession: B59151; A59151
R;Oliveira, A.E.A.; Machado, O.L.T.; Gomes, V.M.; Xavier-Neto, J.; Pereira, A.C.P.; Vieira, J.G.H.; Fernandes, K.V.S.; Xavier-Filho, J.
Protein Pept. Lett. 6, 15-21, 1999
A;Title: Jack bean seed coat contains a protein with complete sequence homology to bovine insulin.
A;Reference number: A59151
A;Accession: B59151
A;Molecule type: protein
A;Residues: 1-30 <MACB>
A;Accession: A59151
A;Molecule type: protein
A;Residues: 31-51 <MACA>
C;Comment: The two chains are probably produced from the same precursor.
C;Superfamily: insulin
F;1-30,31-51/Product: insulin #status experimental <MAT>
F;1-30/Domain: chain B #status experimental <CHB>
F;31-51/Domain: chain A #status experimental <CHA>
F;7-37,19-50,36-41/Disulfide bonds: #status predicted

```

Query Match      91.0%;  Score 267.5;  DB 2;  Length 31;
Best Local Similarity  92.3%;  Pred. No. 7.2e-24;
Matches 48;  Conservative  1;  Mismatches  2;  Indels  1;  Gaps  1;

y  1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
   ||||||| ||||| ||||| ||||| ||||| ||| :||||| |||||
b  1 FVNQHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCTASVCSLYQLENYCN 51

```

RESULT 8

IPHU
insulin precursor [validated] - human
N;Alternate names: preproinsulin
C;Species: Homo sapiens (man)
C;Date: 23-Oct-1981 #sequence_revision 23-Oct-1981 #text_change 08-Dec-2000
C;Accession: A93222; A94253; A93216; A94251; A93144; A92075; A91186; I58114;
A01579; S58661
R;Bell, G.I.; Pictet, R.L.; Rutter, W.J.; Cordell, B.; Tischer, E.; Goodman,
H.M.
Nature 284, 26-32, 1980
A;Title: Sequence of the human insulin gene.
A;Reference number: A93222; MUID:80120725; PMID:6243748
A;Accession: A93222
A;Molecule type: DNA
A;Residues: 1-110 <BEL>
A;Cross-references: GB:J00265; NID:g186429; PIDN:AAA59172.1; PID:g386828
R;Ullrich, A.; Dull, T.J.; Gray, A.; Brosius, J.; Sures, I.
Science 209, 612-615, 1980
A;Title: Genetic variation in the human insulin gene.
A;Reference number: A94253; MUID:80236313; PMID:6248962
A;Accession: A94253
A;Molecule type: DNA
A;Residues: 1-110 <ULL>
A;Cross-references: GB:J00265; NID:g186429; PIDN:AAA59172.1; PID:g386828

R;Bell, G.I.; Swain, W.F.; Pictet, R.; Cordell, B.; Goodman, H.M.; Rutter, W.J.
Nature 282, 525-527, 1979
A;Title: Nucleotide sequence of a cDNA clone encoding human preproinsulin.
A;Reference number: A93216; MUID:80054779; PMID:503234
A;Accession: A93216
A;Molecule type: mRNA
A;Residues: 1-110 <BEL2>
A;Cross-references: GB:J00265; NID:g186429; PIDN:AAA59172.1; PID:g386828
R;Sures, I.; Goeddel, D.V.; Gray, A.; Ullrich, A.
Science 208, 57-59, 1980
A;Title: Nucleotide sequence of human preproinsulin complementary DNA.
A;Reference number: A94251; MUID:80147417; PMID:6927840
A;Accession: A94251
A;Molecule type: mRNA
A;Residues: 1-110 <SUR>
A;Cross-references: GB:J00265; NID:g186429; PIDN:AAA59172.1; PID:g386828
R;Nicol, D.S.H.W.; Smith, L.F.
Nature 187, 483-485, 1960
A;Title: Amino-acid sequence of human insulin.
A;Reference number: A93144
A;Accession: A93144
A;Molecule type: protein
A;Residues: 25-54;90-110 <NIC>
R;Oyer, P.E.; Cho, S.; Peterson, J.D.; Steiner, D.F.
J. Biol. Chem. 246, 1375-1386, 1971
A;Title: Studies on human proinsulin. Isolation and amino acid sequence of the
human pancreatic C-peptide.
A;Reference number: A92075; MUID:71116410; PMID:5101771
A;Accession: A92075
A;Molecule type: protein
A;Residues: 57-87 <OYE>
R;Ko, A.; Smyth, D.G.; Markussen, J.; Sundby, F.
Eur. J. Biochem. 20, 190-199, 1971
A;Title: Amino acid sequence of the C-peptide of human proinsulin.
A;Reference number: A91186; MUID:71257722; PMID:5560404
A;Accession: A91186
A;Molecule type: protein
A;Residues: 57-87 <KOA>
R;Lucassen, A.M.; Julier, C.; Beressi, J.P.; Boitard, C.; Froguel, P.; Lathrop,
M.; Bell, J.I.
Nature Genet. 4, 305-310, 1993
A;Title: Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb
segment of DNA spanning the insulin gene and associated VNTR.
A;Reference number: I58114; MUID:93364428; PMID:8358440
A;Accession: I58114
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-59,63-110 <RES>
A;Cross-references: GB:L15440; NID:g307071; PIDN:AAA59179.1; PID:g307072
R;Sieber, P.; Kamber, B.; Hartmann, A.; Joehl, A.; Riniker, B.; Rittel, W.
Helv. Chim. Acta 57, 2617-2621, 1974
A;Title: Totalsynthese von Humaninsulin unter gezielter Bildung der
Disulfidbindungen.
A;Reference number: A91636; MUID:75077277; PMID:4443293
A;Contents: annotation; synthesis
A;Note: disulfide-bonded human insulin was synthesized; the synthetic hormone
was identical with the natural hormone in chemical and biological activities

A;Note: article in German with English abstract
 R;Naithani, V.K.
 Hoppe-Seyler's Z. Physiol. Chem. 354, 659-672, 1973
 A;Title: The synthesis of C-peptide of human proinsulin.
 A;Reference number: A91658; MUID:75040007; PMID:4803504
 A;Contents: annotation; synthesis of residues 57-87
 R;Geiger, R.; Jaeger, G.; Koenig, W.
 Chem. Ber. 106, 2347-2352, 1973
 A;Title: Synthesis of the complete sequence of human proinsulin C-peptide and its [Glu-9,Gln-11] analogue.
 A;Reference number: A90914
 A;Contents: annotation; synthesis of residues 57-87
 R;Kaufmann, J.E.; Irminger, J.C.; Halban, P.A.
 Biochem. J. 310, 869-874, 1995
 A;Title: Sequence requirements for proinsulin processing at the B-chain/C-peptide junction.
 A;Reference number: S58661; MUID:96013185; PMID:7575420
 A;Contents: annotation; site-directed mutagenesis study of proteolytic processing
 C;Genetics:
 A;Gene: GDB:INS
 A;Cross-references: GDB:119349; OMIM:176730
 A;Map position: 11p15.5-11p15.5
 A;Introns: 63/1
 C;Superfamily: insulin
 C;Keywords: hormone; pancreas
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;25-54/Domain: insulin chain B #status experimental <BCB>
 F;25-54,90-110/Product: insulin #status experimental <MAT>
 F;57-87/Domain: connecting C peptide #status experimental <CPEP>
 F;90-110/Domain: insulin chain A #status experimental <ACH>
 F;31-96,43-109,95-100/Disulfide bonds: #status experimental

 Query Match 90.8%; Score 267; DB 1; Length 110;
 Best Local Similarity 60.5%; Pred. No. 1.6e-23;
 Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

 Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 30
 Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 84

 Qy 31 ----RGIVEQCCTSICSLYQLENYCN 52
 Db 85 SLQKRGIVEQCCTSICSLYQLENYCN 110

RESULT 9
 B42179
 insulin precursor - green monkey
 C;Species: Cercopithecus aethiops (green monkey, grivet)
 C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999
 C;Accession: B42179; A05232; S16494; S22056
 R;Seino, S.; Bell, G.I.; Li, W.H.
 Mol. Biol. Evol. 9, 193-203, 1992
 A;Title: Sequences of primate insulin genes support the hypothesis of a slower rate of molecular evolution in humans and apes than in monkeys.
 A;Reference number: A42179; MUID:92219953; PMID:1560757

A;Accession: B42179
 A;Molecule type: DNA
 A;Residues: 1-110 <SEI>
 A;Cross-references: EMBL:X61092; NID:g22808; PIDN:CAA43405.1; PID:g22809
 A;Note: sequence extracted from NCBI backbone (NCBIN:95185, NCBIP:95194)
 R;Peterson, J.D.; Nehrlich, S.; Oyer, P.E.; Steiner, D.F.
J. Biol. Chem. 247, 4866-4871, 1972
 A;Title: Determination of the amino acid sequence of the monkey, sheep, and dog
 proinsulin C-peptides by a semi-micro Edman degradation procedure.
 A;Reference number: A92111; MUID:72258016; PMID:4626369
 A;Accession: A05232
 A;Molecule type: protein
 A;Residues: 57-87 <PET>
 C;Genetics:
 A;Introns: 63/1
 C;Superfamily: insulin
 C;Keywords: hormone; pancreas
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;25-54/Domain: insulin chain B #status predicted <BCH>
 F;25-54,90-110/Product: insulin #status predicted <MAT>
 F;57-87/Domain: connecting peptide #status experimental <CPEP>
 F;90-110/Domain: insulin chain A #status predicted <ACH>
 F;31-96,43-109,95-100/Disulfide bonds: #status predicted

 Query Match 90.8%; Score 267; DB 2; Length 110;
 Best Local Similarity 60.5%; Pred. No. 1.6e-23;
 Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

 QY 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 30
 | ||||||| ||||| ||||| ||||| ||||| |||||
 Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGPGAGSLQPLALEG 84

 QY 31 ----RGIVEQCCTSICSLYQLENYCN 52
 | ||||||| ||||| ||||| |||||
 Db 85 SLQKRGIVEQCCTSICSLYQLENYCN 110

RESULT 10

A42179

insulin precursor - chimpanzee

C;Species: Pan troglodytes (chimpanzee)

C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999

C;Accession: A42179; S22058

R;Seino, S.; Bell, G.I.; Li, W.H.

Mol. Biol. Evol. 9, 193-203, 1992

A;Title: Sequences of primate insulin genes support the hypothesis of a slower
 rate of molecular evolution in humans and apes than in monkeys.

A;Reference number: A42179; MUID:92219953; PMID:1560757

A;Accession: A42179

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-110 <SEI>

A;Cross-references: EMBL:X61089; NID:g38251; PIDN:CAA43403.1; PID:g38252

A;Note: sequence extracted from NCBI backbone (NCBIP:95067)

C;Genetics:

A;Introns: 63/1

C;Superfamily: insulin

Query Match 90.8%; Score 267; DB 2; Length 110;
Best Local Similarity 60.5%; Pred. No. 1.6e-23;
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 30
| | | | | | | | | | | | | | | | | | | | | | | |
Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 84

Qy 31 ----RGIVEQCCTSICSLYQLENYCN 52
| | | | | | | | | | | | | | | | | | | | | | | |
Db 85 SLQKRGIVEQCCTSICSLYQLENYCN 110

RESULT 11

JQ0178

insulin precursor - crab-eating macaque
C;Species: Macaca fascicularis (crab-eating macaque)
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 16-Jul-1999
C;Accession: JQ0178
R;Wetekam, W.; Groneberg, J.; Leineweber, M.; Wengenmayer, F.; Winnacker, E.L.
Gene 19, 179-183, 1982
A;Title: The nucleotide sequence of cDNA coding for preproinsulin from the
primate Macaca fascicularis.
A;Reference number: JQ0178; MUID:83080474; PMID:6184262
A;Accession: JQ0178
A;Molecule type: mRNA
A;Residues: 1-110 <WET>
A;Cross-references: GB:J00336; NID:g342121; PIDN:AAA36849.1; PID:g342122
C;Superfamily: insulin
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-54,90-110/Product: insulin #status predicted <MAT>
F;25-54/Domain: insulin chain B #status predicted <BCH>
F;55-89/Domain: insulin connecting C peptide #status predicted <CPT>
F;90-110/Domain: insulin chain A #status predicted <ACH>
F;31-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 90.8%; Score 267; DB 2; Length 110;
Best Local Similarity 60.5%; Pred. No. 1.6e-23;
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 30
| | | | | | | | | | | | | | | | | | | | | | | |
Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGPGAGSLQPLALEG 84

Qy 31 ----RGIVEQCCTSICSLYQLENYCN 52
| | | | | | | | | | | | | | | | | | | | | | | |
Db 85 SLQKRGIVEQCCTSICSLYQLENYCN 110

RESULT 12

INWH1S

insulin - sei whale

C;Species: Balaenoptera borealis (sei whale)
C;Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 16-Jul-1999
C;Accession: A01582
R;Ishihara, Y.; Saito, T.; Ito, Y.; Fujino, M.

Nature 181, 1468-1469, 1958
A;Title: Structure of sperm- and sei-whale insulins and their breakdown by whale pepsin.
A;Reference number: A93142
A;Accession: A01582
A;Molecule type: protein
A;Residues: 1-30;31-51 <ISH>
C;Superfamily: insulin
C;Keywords: hormone; pancreas
F;1-30/Domain: insulin chain B #status experimental <BCN>
F;1-30,31-51/Product: insulin #status experimental <MAT>
F;31-51/Domain: insulin chain A #status experimental <ACH>
F;7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 89.6%; Score 263.5; DB 1; Length 51;
Best Local Similarity 92.3%; Pred. No. 2.1e-23;
Matches 48; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
| || || || | || || || | || || || | || || || | || || || | || || || |
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCASTCSLYQLENYCN 51

RESULT 13

INGT
insulin - goat
C;Species: Capra aegagrus hircus (domestic goat)
C;Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 16-Jul-1999
C;Accession: A01586
R;Smith, L.F.
Am. J. Med. 40, 662-666, 1966
A;Title: Species variation in the amino acid sequence of insulin.
A;Reference number: A90029; MUID:66160119; PMID:5949593
A;Accession: A01586
A;Molecule type: protein
A;Residues: 1-30;31-51 <SMI>
C;Superfamily: insulin
C;Keywords: hormone; pancreas
F;1-30/Domain: insulin chain B #status experimental <BCN>
F;1-30,31-51/Product: insulin #status experimental <MAT>
F;31-51/Domain: insulin chain A #status experimental <ACH>
F;7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 89.6%; Score 263.5; DB 1; Length 51;
Best Local Similarity 90.4%; Pred. No. 2.1e-23;
Matches 47; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
| || || || | || || || | || || || | || || || | :|| || || || |
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCAGVCSLYQLENYCN 51

RESULT 14

INCMA
insulin - Arabian camel (tentative sequence)
C;Species: Camelus dromedarius (Arabian camel)
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999

A;Residues: 33-38,40-62 <SNE>
A;Note: the authors report the characterization of a connecting peptide variant lacking Ala-39
A;Accession: B60835
A;Molecule type: protein
A;Residues: 33-62 <SN2>
R;Blundell, T.; Dodson, G.; Hodgkin, D.; Mercola, D.
Adv. Protein Chem. 26, 279-402, 1972
A;Title: Insulin. the structure in the crystal and its reflection in chemistry and biology.
A;Reference number: A90017
A;Contents: annotation; X-ray crystallography, 1.9 angstroms
C;Superfamily: insulin
C;Keywords: hormone; pancreas
F;1-30/Domain: insulin chain B #status experimental <BCN>
F;1-30,64-84/Product: insulin #status experimental <MAT>
F;33-63/Domain: connecting peptide #status experimental <CPPEP>
F;64-84/Domain: insulin chain A #status experimental <ACH>
F;7-70,19-83,69-74/Disulfide bonds: #status experimental

Query Match 89.5%; Score 263; DB 1; Length 84;
Best Local Similarity 60.7%; Pred. No. 3.6e-23;
Matches 51; Conservative 0; Mismatches 1; Indels 32; Gaps 1;

Qy	1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT-----	30
Db	1 FVNQHLCGSHLVEALYLVCGERGFFYTPKARREAENPQAGAVELGGGLGGLQALALEGPP	60
Qy	31 --RGIVEQCCTSICSLYQLENYCN 52	
Db	61 QKRGIVEQCCTSICSLYQLENYCN 84	

Search completed: July 15, 2004, 16:37:33
Job time : 5.98756 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 16:37:41 ; Search time 21.6343 Seconds
(without alignments)
751.267 Million cell updates/sec

Title: US-09-423-100-5
Perfect score: 294
Sequence: 1 FVNQHLCGSHLVEALYLVCG.....IVEQCCTSICSLYQLENYCN 52

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1285345 seqs, 312560633 residues

Total number of hits satisfying chosen parameters: 1285345

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

%
Result Query
No. Score Match Length DB ID Description

1	294	100.0	52	13	US-10-054-873-5	Sequence 5, Appli
2	294	100.0	107	13	US-10-054-873-6	Sequence 6, Appli
3	294	100.0	137	16	US-10-101-454-39	Sequence 39, Appli
4	294	100.0	145	16	US-10-101-454-45	Sequence 45, Appli
5	294	100.0	146	16	US-10-101-454-48	Sequence 48, Appli
6	294	100.0	150	13	US-10-054-873-7	Sequence 7, Appli
7	278.5	94.7	51	10	US-09-858-935B-5	Sequence 5, Appli
8	278.5	94.7	51	12	US-10-444-649-3	Sequence 3, Appli
9	278.5	94.7	51	12	US-10-444-701-3	Sequence 3, Appli
10	278.5	94.7	51	12	US-10-271-869-5	Sequence 5, Appli
11	278.5	94.7	51	13	US-10-028-410-3	Sequence 3, Appli
12	278.5	94.7	51	14	US-10-444-326-3	Sequence 3, Appli
13	278.5	94.7	51	16	US-10-444-262-3	Sequence 3, Appli
14	275.5	93.7	104	16	US-10-101-454-15	Sequence 15, Appli
15	275.5	93.7	124	9	US-09-894-711-18	Sequence 18, Appli
16	275.5	93.7	138	9	US-09-861-687-19	Sequence 19, Appli
17	275.5	93.7	138	12	US-10-620-651-19	Sequence 19, Appli
18	275.5	93.7	140	16	US-10-101-454-33	Sequence 33, Appli
19	275.5	93.7	140	16	US-10-101-454-42	Sequence 42, Appli
20	273	92.9	50	13	US-10-066-009A-3	Sequence 3, Appli
21	271.5	92.3	102	16	US-10-101-454-36	Sequence 36, Appli
22	267	90.8	86	9	US-09-878-380-1	Sequence 1, Appli
23	267	90.8	86	10	US-09-858-935B-4	Sequence 4, Appli
24	267	90.8	86	12	US-10-444-649-2	Sequence 2, Appli
25	267	90.8	86	12	US-10-444-701-2	Sequence 2, Appli
26	267	90.8	86	12	US-10-271-869-4	Sequence 4, Appli
27	267	90.8	86	13	US-10-028-410-2	Sequence 2, Appli
28	267	90.8	86	13	US-10-054-873-4	Sequence 4, Appli
29	267	90.8	86	14	US-10-444-326-2	Sequence 2, Appli
30	267	90.8	86	16	US-10-444-262-2	Sequence 2, Appli
31	267	90.8	96	9	US-09-947-563-4	Sequence 4, Appli
32	267	90.8	110	9	US-09-205-658-125	Sequence 125, App
33	267	90.8	110	9	US-09-815-229-3	Sequence 3, Appli
34	267	90.8	110	9	US-09-804-409A-9	Sequence 9, Appli
35	267	90.8	110	10	US-09-969-748C-6	Sequence 6, Appli
36	267	90.8	110	10	US-09-963-693-125	Sequence 125, App
37	267	90.8	110	12	US-10-411-037-44	Sequence 44, Appli
38	267	90.8	110	12	US-10-411-026-44	Sequence 44, Appli
39	267	90.8	110	14	US-10-038-686-1	Sequence 1, Appli
40	267	90.8	110	14	US-10-328-813-2	Sequence 2, Appli
41	267	90.8	110	15	US-10-383-285-2	Sequence 2, Appli
42	267	90.8	110	15	US-10-346-563-2	Sequence 2, Appli
43	267	90.8	110	15	US-10-321-717-2	Sequence 2, Appli
44	267	90.8	110	16	US-10-410-962-44	Sequence 44, Appli
45	267	90.8	110	16	US-10-411-049-44	Sequence 44, Appli

ALIGNMENTS

```

RESULT 1
US-10-054-873-5
; Sequence 5, Application US/10054873
; Publication No. US20020164712A1
; GENERAL INFORMATION:
;      APPLICANT: Gan, Zhong Ru

```

; TITLE OF INVENTION: Chimeric Protein Containing an
; Intramolecular Chaperone-Like Sequence
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/054,873
; FILING DATE: 22-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/CN98/00052
; FILING DATE: 31-MAR-1998
; APPLICATION NUMBER: US 09/423,100
; FILING DATE: 11-DEC-2000
; ATTORNEY/AGENT INFORMATION:
; NAME: Mycroft, Frank J
; REGISTRATION NUMBER: 46,946
; REFERENCE/DOCKET NUMBER: 020167-000130US
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 52 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-054-873-5

Query Match 100.0%; Score 294; DB 13; Length 52;
Best Local Similarity 100.0%; Pred. No. 5e-29;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52

RESULT 2
US-10-054-873-6
; Sequence 6, Application US/10054873
; Publication No. US20020164712A1
; GENERAL INFORMATION:
; APPLICANT: Gan, Zhong Ru
; TITLE OF INVENTION: Chimeric Protein Containing an
; Intramolecular Chaperone-Like Sequence
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/054,873
; FILING DATE: 22-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/CN98/00052
; FILING DATE: 31-MAR-1998
; APPLICATION NUMBER: US 09/423,100
; FILING DATE: 11-DEC-2000
; ATTORNEY/AGENT INFORMATION:
; NAME: Mycroft, Frank J
; REGISTRATION NUMBER: 46,946
; REFERENCE/DOCKET NUMBER: 020167-000130US
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 107 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-054-873-6

Query Match 100.0%; Score 294; DB 13; Length 107;
Best Local Similarity 100.0%; Pred. No. 1.1e-28;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 56 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107

RESULT 3
US-10-101-454-39
; Sequence 39, Application US/10101454
; Publication No. US20040110664A1
; GENERAL INFORMATION:
; APPLICANT: Havelund, Svend
; Halstrom, John
; Jonassen, Ib
; Andersen, Asser Sloth
; Markussen, Jan
; TITLE OF INVENTION: ACYLATED INSULIN
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Novo Nordisk of North America, Inc.

;
; STREET: 405 Lexington Avenue, 64th Floor
;
; CITY: New York
;
; STATE: New York
;
; COUNTRY: United States of America
;
; ZIP: 10174-6401
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/10/101,454
;
; FILING DATE: 20-Mar-2002
;
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER: 08/400,256
;
; FILING DATE: 03-MAR-1995
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: Lambiris, Elias J.
;
; REGISTRATION NUMBER: 33,728
;
; REFERENCE/DOCKET NUMBER: 3985.220-US
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: 212-867-0123
;
; TELEFAX: 212-878-9655
;
; INFORMATION FOR SEQ ID NO: 39:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 137 amino acids
;
; TYPE: amino acid
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-10-101-454-39

Query Match 100.0%; Score 294; DB 16; Length 137;
Best Local Similarity 100.0%; Pred. No. 1.4e-28;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 86 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 137

RESULT 4
US-10-101-454-45
; Sequence 45, Application US/10101454
; Publication No. US20040110664A1
; GENERAL INFORMATION:
; APPLICANT: Havelund, Svend
; Halstrom, John
; Jonassen, Ib
; Andersen, Asser Sloth
; Markussen, Jan
; TITLE OF INVENTION: ACYLATED INSULIN
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Novo Nordisk of North America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor

;
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
;
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/101,454
; FILING DATE: 20-Mar-2002
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/400,256
; FILING DATE: 03-MAR-1995
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3985.220-US
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
;
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 145 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 45:
US-10-101-454-45

Query Match 100.0%; Score 294; DB 16; Length 145;
Best Local Similarity 100.0%; Pred. No. 1.5e-28;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 145

RESULT 5
US-10-101-454-48
; Sequence 48, Application US/10101454
; Publication No. US20040110664A1
; GENERAL INFORMATION:
; APPLICANT: Havelund, Svend
; Halstrom, John
; Jonassen, Ib
; Andersen, Asser Sloth
; Markussen, Jan
; TITLE OF INVENTION: ACYLATED INSULIN
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Novo Nordisk of North America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York

; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/101,454
; FILING DATE: 20-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/400,256
; FILING DATE: 03-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3985.220-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 146 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 48:
US-10-101-454-48

Query Match 100.0%; Score 294; DB 16; Length 146;
Best Local Similarity 100.0%; Pred. No. 1.5e-28;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 95 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 146

RESULT 6
US-10-054-873-7
; Sequence 7, Application US/10054873
; Publication No. US20020164712A1
; GENERAL INFORMATION:
; APPLICANT: Gan, Zhong Ru
; TITLE OF INVENTION: Chimeric Protein Containing an
; Intramolecular Chaperone-Like Sequence
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/054,873
; FILING DATE: 22-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/CN98/00052
; FILING DATE: 31-MAR-1998
; APPLICATION NUMBER: US 09/423,100
; FILING DATE: 11-DEC-2000
; ATTORNEY/AGENT INFORMATION:
; NAME: Mycroft, Frank J
; REGISTRATION NUMBER: 46,946
; REFERENCE/DOCKET NUMBER: 020167-000130US
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 150 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-10-054-873-7

Query Match 100.0%; Score 294; DB 13; Length 150;
Best Local Similarity 100.0%; Pred. No. 1.6e-28;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 99 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 150

RESULT 7
US-09-858-935B-5
; Sequence 5, Application US/09858935B
; Publication No. US20030069177A1
; GENERAL INFORMATION:
; APPLICANT: Dubaquie, Yves
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Lowman, Henry B.
; TITLE OF INVENTION: METHOD FOR TREATING CARTILAGE DISORDERS
; FILE REFERENCE: P1794R1
; CURRENT APPLICATION NUMBER: US/09/858,935B
; CURRENT FILING DATE: 2002-07-02
; PRIOR APPLICATION NUMBER: US 60/248,985
; PRIOR FILING DATE: 2000-11-15
; PRIOR APPLICATION NUMBER: US 60/204,490
; PRIOR FILING DATE: 2000-05-16
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 5
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-858-935B-5

Query Match 94.7%; Score 278.5; DB 10; Length 51;
Best Local Similarity 98.1%; Pred. No. 4.1e-27;
Matches 51; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT-GIVEQCCTSICSLYQLENYCN 51

RESULT 8

US-10-444-649-3

; Sequence 3, Application US/10444649
; Publication No. US20040033951A1
; GENERAL INFORMATION:
; APPLICANT: Dubaquie, Yves
; APPLICANT: Lowman, Henry
; TITLE OF INVENTION: PROTEIN VARIANTS
; FILE REFERENCE: P1712R1
; CURRENT APPLICATION NUMBER: US/10/444,649
; CURRENT FILING DATE: 2003-05-22
; PRIOR APPLICATION NUMBER: US/09/724,479
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: US/09/477,923
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 3
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens

US-10-444-649-3

Query Match 94.7%; Score 278.5; DB 12; Length 51;
Best Local Similarity 98.1%; Pred. No. 4.1e-27;
Matches 51; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT-GIVEQCCTSICSLYQLENYCN 51

RESULT 9

US-10-444-701-3

; Sequence 3, Application US/10444701
; Publication No. US20040033952A1
; GENERAL INFORMATION:
; APPLICANT: Dubaquie, Yves
; APPLICANT: Lowman, Henry
; TITLE OF INVENTION: PROTEIN VARIANTS
; FILE REFERENCE: P1712R1
; CURRENT APPLICATION NUMBER: US/10/444,701
; CURRENT FILING DATE: 2003-05-22
; PRIOR APPLICATION NUMBER: US/09/723,866
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: US/09/477,923
; PRIOR FILING DATE: 2000-01-05

; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 3
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-444-701-3

Query Match 94.7%; Score 278.5; DB 12; Length 51;
Best Local Similarity 98.1%; Pred. No. 4.1e-27;
Matches 51; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT-GIVEQCCTSICSLYQLENYCN 51

RESULT 10

US-10-271-869-5

; Sequence 5, Application US/10271869
; Publication No. US20030211992A1
; GENERAL INFORMATION:
; APPLICANT: Dubaquie, Yves
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Lowman, Henry B.
; TITLE OF INVENTION: METHOD FOR TREATING CARTILAGE DISORDERS
; FILE REFERENCE: P1794R1
; CURRENT APPLICATION NUMBER: US/10/271,869
; CURRENT FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US/09/858,935
; PRIOR FILING DATE: 2002-07-02
; PRIOR APPLICATION NUMBER: US 60/248,985
; PRIOR FILING DATE: 2000-11-15
; PRIOR APPLICATION NUMBER: US 60/204,490
; PRIOR FILING DATE: 2000-05-16
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 5
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-271-869-5

Query Match 94.7%; Score 278.5; DB 12; Length 51;
Best Local Similarity 98.1%; Pred. No. 4.1e-27;
Matches 51; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT-GIVEQCCTSICSLYQLENYCN 51

RESULT 11

US-10-028-410-3

; Sequence 3, Application US/10028410
; Publication No. US20020160955A1
; GENERAL INFORMATION:
; APPLICANT: Dubaquie, Yves
; APPLICANT: Lowman, Henry

; TITLE OF INVENTION: PROTEIN VARIANTS
; FILE REFERENCE: P1712R1-1
; CURRENT APPLICATION NUMBER: US/10/028,410
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US/09/477,924
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 3
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-028-410-3

Query Match 94.7%; Score 278.5; DB 13; Length 51;
Best Local Similarity 98.1%; Pred. No. 4.1e-27;
Matches 51; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT-GIVEQCCTSICSLYQLENYCN 51

RESULT 12
US-10-444-326-3
; Sequence 3, Application US/10444326
; Publication No. US20030191065A1
; GENERAL INFORMATION:
; APPLICANT: Dubaquie, Yves
; APPLICANT: Lowman, Henry
; TITLE OF INVENTION: PROTEIN VARIANTS
; FILE REFERENCE: P1712R1
; CURRENT APPLICATION NUMBER: US/10/444,326
; CURRENT FILING DATE: 2003-05-22
; PRIOR APPLICATION NUMBER: US/09/723,866
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: US/09/477,923
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 3
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-444-326-3

Query Match 94.7%; Score 278.5; DB 14; Length 51;
Best Local Similarity 98.1%; Pred. No. 4.1e-27;
Matches 51; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT-GIVEQCCTSICSLYQLENYCN 51

RESULT 13
US-10-444-262-3
; Sequence 3, Application US/10444262
; Publication No. US20040023883A1

; GENERAL INFORMATION:
; APPLICANT: Dubaque, Yves
; APPLICANT: Lowman, Henry
; TITLE OF INVENTION: PROTEIN VARIANTS
; FILE REFERENCE: P1712R1
; CURRENT APPLICATION NUMBER: US/10/444,262
; CURRENT FILING DATE: 2003-05-22
; PRIOR APPLICATION NUMBER: US/09/724,478
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: US/09/477,923
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 3
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-444-262-3

Query Match 94.7%; Score 278.5; DB 16; Length 51;
Best Local Similarity 98.1%; Pred. No. 4.1e-27;
Matches 51; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT-GIVEQCCTSICSLYQLENYCN 51

RESULT 14
US-10-101-454-15
; Sequence 15, Application US/10101454
; Publication No. US20040110664A1
; GENERAL INFORMATION:
; APPLICANT: Havelund, Svend
; Halstrom, John
; Jonassen, Ib
; Andersen, Asser Sloth
; Markussen, Jan
; TITLE OF INVENTION: ACYLATED INSULIN
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Novo Nordisk of North America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/101,454
; FILING DATE: 20-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/400,256

; FILING DATE: 03-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3985.220-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 104 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-10-101-454-15

Query Match 93.7%; Score 275.5; DB 16; Length 104;
Best Local Similarity 90.9%; Pred. No. 2.1e-26;
Matches 50; Conservative 2; Mismatches 0; Indels 3; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT---RGIVEQCCTSICSLYQLENYCN 52
|||:|||||:|||||:|||||:|||||: :|||||:|||||:|||||:
Db 50 FVNQHLCGSHLVEALYLVCGERGFFYTPKSDDAKGIVEQCCTSICSLYQLENYCN 104

RESULT 15
US-09-894-711-18
; Sequence 18, Application US/09894711
; Patent No. US20020137144A1
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas Borglum
; APPLICANT: Ludvigsen, Svend
; TITLE OF INVENTION: Method for making insulin precursors and
; TITLE OF INVENTION: insulin precursor analogues having improved fermentation
; TITLE OF INVENTION: yield in yeast
; FILE REFERENCE: 6148.400-US
; CURRENT APPLICATION NUMBER: US/09/894,711
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: PA 2000 00443
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: PA 1999 01869
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/211,081
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/181,450
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: 09/740,359
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic

US-09-894-711-18

Query Match 93.7%; Score 275.5; DB 9; Length 124;
Best Local Similarity 94.3%; Pred. No. 2.5e-26;
Matches 50; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPK-TRGIVEQCCTSICSLYQLENYCN 52
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 72 FVNQHLCGSHLVEALYLVCGERGFFYTPKAAGIVEQCCTSICSLYQLENYCN 124

Search completed: July 15, 2004, 17:05:08
Job time : 21.6343 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 16:29:50 ; Search time 17.7537 Seconds
(without alignments)
924.141 Million cell updates/sec

Title: US-09-423-100-5

Perfect score: 294

Sequence: 1 FVNQHLCGSHLVEALYLVCG.....IVEQCCTSICSLYQLENYCN 52

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:
1: sp_archea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phage:
10: sp_plant:
11: sp_rat:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_rvirus:
16: sp_bacteriap:
17: sp_archeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			Description
		Match	Length	DB ID	

1	267	90.8	110	6	Q8HXV2	Q8hxv2 pongo pygma
2	251	85.4	110	6	Q8WNW6	Q8wnw6 felis silve
3	219.5	74.7	106	13	Q9I8Q7	Q9i8q7 rana pipien
4	201.5	68.5	110	13	Q98TA8	Q98ta8 pantodon bu
5	195.5	66.5	108	13	Q9DDE5	Q9dde5 brachydanio
6	195.5	66.5	108	13	Q90ZN4	Q90zn4 catla catla
7	195	66.3	111	13	Q98TB0	Q98tb0 chitala chi
8	193.5	65.8	110	13	Q90ZY1	Q90zy1 hiodon alos
9	191.5	65.1	111	13	Q98TA7	Q98ta7 osteoglossu
10	187.5	63.8	87	13	Q98TA9	Q98ta9 gnathonemus
11	186.5	63.4	108	13	Q98TB1	Q98tbl1 catostomus
12	185.5	63.1	91	13	Q98TB2	Q98tb2 ambloplites
13	146	49.7	65	6	Q8HZ81	Q8hz81 gorilla gor
14	146	49.7	65	6	Q8HZ80	Q8hz80 pongo pygma
15	144	49.0	207	13	Q90XD0	Q90xd0 cyprinus ca
16	144	49.0	215	13	O73721	O73721 tilapia sp.
17	143.5	48.8	132	13	Q8AV14	Q8av14 petromyzon
18	141.5	48.1	104	13	Q7T107	Q7t107 dicentrarch
19	141.5	48.1	108	13	Q800N0	Q800n0 morone chry
20	141.5	48.1	108	13	Q800M9	Q800m9 morone saxa
21	141.5	48.1	108	13	Q800M8	Q800m8 morone chry
22	141.5	48.1	108	13	Q800M7	Q800m7 morone amer
23	141.5	48.1	159	13	O93607	O93607 paralichthy
24	141.5	48.1	182	13	O73720	O73720 oreochromis
25	141.5	48.1	182	13	O42289	O42289 oreochromis
26	141.5	48.1	182	13	P79824	P79824 oreochromis
27	141.5	48.1	185	13	O57436	O57436 paralichthy
28	141.5	48.1	186	13	O93527	O93527 paralichthy
29	141.5	48.1	186	13	Q800Y5	Q800y5 siganus gut
30	141.5	48.1	186	13	Q7T1A7	Q7t1a7 perca flave
31	141	48.0	185	13	Q9YI57	Q9yi57 acanthopagr
32	140.5	47.8	116	13	Q91161	Q91161 oncorhynchus
33	140.5	47.8	117	13	Q91476	Q91476 salmo salar
34	140.5	47.8	145	13	Q91475	Q91475 salmo salar
35	140.5	47.8	149	13	Q91231	Q91231 oncorhynchus
36	140.5	47.8	155	13	Q91162	Q91162 oncorhynchus
37	140.5	47.8	161	13	Q91230	Q91230 oncorhynchus
38	140.5	47.8	184	13	O42336	O42336 myoxocephal
39	140.5	47.8	188	13	P81268	P81268 oncorhynchus
40	140.5	47.8	188	13	Q91965	Q91965 oncorhynchus
41	140	47.6	210	13	Q91443	Q91443 squalus aca
42	140	47.6	215	13	Q800Y4	Q800y4 siganus gut
43	139	47.3	212	13	Q8JIE4	Q8jie4 brachydanio
44	139	47.3	215	13	O42429	O42429 lates calca
45	139	47.3	215	13	Q800E6	Q800e6 paralichthy

ALIGNMENTS

RESULT 1

Q8HXV2

ID Q8HXV2 PRELIMINARY; PRT; 110 AA.
 AC Q8HXV2;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Insulin precursor.
 GN INS.
 OS Pongo pygmaeus (Orangutan).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
 OX NCBI_TaxID=9600;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Stead J.D.H., Jeffreys A.J.;
 RT "Haplotype diversity at the insulin region.";
 RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AY137503; AAN06937.1; -.
 DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR GO; GO:0007582; P:physiological processes; IEA.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; I1GF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 SQ SEQUENCE 110 AA; 12038 MW; 22D2B32B94F520F8 CRC64;

 Query Match 90.8%; Score 267; DB 6; Length 110;
 Best Local Similarity 60.5%; Pred. No. 7.8e-29;
 Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

 Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 30
 |||||||
 Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 84

 Qy 31 ----RGIVEQCCTSICSLYQLENYCN 52
 |||||||
 Db 85 SLQKRGIVEQCCTSICSLYQLENYCN 110

RESULT 2
 Q8WNW6
 ID Q8WNW6 PRELIMINARY; PRT; 110 AA.
 AC Q8WNW6;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Preproinsulin.
 OS Felis silvestris catus (Cat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
 OX NCBI TaxID=9685;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RA Okamoto S., Morimatsu M.;
 RT "cat insulin.";
 RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 DR EMBL; AB043535; BAB84110.1; -.
 DR GO; GO:0005576; C:extracellular; IEA.

DR GO; GO:0005179; F:hormone activity; IEA.
DR GO; GO:0007582; P:physiological processes; IEA.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; I1GF; 1.
DR PROSITE; PS00262; INSULIN; 1.
SQ SEQUENCE 110 AA; 12069 MW; 95FB6E170C7BECA4 CRC64;

Query Match 85.4%; Score 251; DB 6; Length 110;
Best Local Similarity 55.8%; Pred. No. 1.2e-26;
Matches 48; Conservative 2; Mismatches 2; Indels 34; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 30
| | | | | | | | | | | | | | | | | | | | | |
Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKARREAEDLQGKDAELGEAPGAGGLQPSALEA 84

Qy 31 ----RGIVEQCCTSICSLYQLENYCN 52
| | | | | | : | : | | | : | |
Db 85 PLQKRGIVEQCCASVCSLYQLEHYCN 110

RESULT 3

Q9I8Q7

ID Q9I8Q7 PRELIMINARY; PRT; 106 AA.
AC Q9I8Q7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Preproinsulin.
OS Rana pipiens (Northern leopard frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Ranidae; Rana.
OX NCBI_TaxID=8404;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20362507; PubMed=10818274;
RA Irwin D.M., Sivarajah P.;
RT "Proinsulin cDNAs from the leopard frog, Rana pipiens: evolution of
RT proinsulin processing.";
RL Comp. Biochem. Physiol. 125B:405-410(2000).
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL; AF227187; AAF87285.1; -.
DR HSSP; P01315; 1SDB.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005179; F:hormone activity; IEA.
DR GO; GO:0007582; P:physiological processes; IEA.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; I1GF; 1.
DR PROSITE; PS00262; INSULIN; 1.
SQ SEQUENCE 106 AA; 12183 MW; 3A870EEC70217F92 CRC64;

Query Match 74.7%; Score 219.5; DB 13; Length 106;
Best Local Similarity 49.4%; Pred. No. 2.5e-22;

Matches 41; Conservative 7; Mismatches 4; Indels 31; Gaps 1;
 Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTR----- 31
 | ||:|||||:|||:||||:|||:
 Db 24 FDNQYLCGSHLVEALYLYMVCGRGFFYSPRSRRDLEQPLVNGLQGSELDEMQVQSQAFQKR 83
 Qy 32 --GIVEQCCTSICSLYQLENYCN 52
 ||||||| : ||||| |||||||
 Db 84 KPGIVEQCCHNTCSLYDLENYCN 106

RESULT 4
Q98TA8
 ID Q98TA8 PRELIMINARY; PRT; 110 AA.
 AC Q98TA8;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Preproinsulin.
 OS Pantodon buchholzii (Butterflyfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;
 OC Osteoglossiformes; Pantodontidae; Pantodon.
 OX NCBI_TaxID=8276;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21203577; PubMed=11306171;
 RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
 RT "Molecular cloning of preproinsulin cDNAs from several
 RT osteoglossomorphs and a cyprinid.";
 RL Mol. Cell. Endocrinol. 174:51-58(2001).
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 DR EMBL; AF199588; AAK28712.1; -.
 DR HSSP; P01308; 1HIS.
 DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR GO; GO:0007582; P:physiological processes; IEA.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; IIIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 SQ SEQUENCE 110 AA; 12324 MW; BDECCD659D872E06 CRC64;
 Query Match 68.5%; Score 201.5; DB 13; Length 110;
 Best Local Similarity 43.5%; Pred. No. 7.7e-20;
 Matches 37; Conservative 8; Mismatches 5; Indels 35; Gaps 1;
 Qy 3 NQHLCGSHLVEALYLVCGERGFFYTPKT----- 30
 :|||||:|||:|||:|||:
 Db 26 SQHLCGSHLVDALYLYMVCGEKGFFYQPKTFRDVDPILLGLSPKSAQENEADEYPYKDQGDL 85
 Qy 31 ---RGIVEQCCTSICSLYQLENYCN 52
 ||||||| |::: |:|||:
 Db 86 KVKGIVEQCCHHPCNIFDLQNYCN 110

RESULT 5
Q9DDE5

ID Q9DDE5 PRELIMINARY; PRT; 108 AA.
AC Q9DDE5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Insulin precursor.
GN INS.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99425190; PubMed=10495291;
RA Argenton F., Zecchin E., Bortolussi M.;
RT "Early appearance of pancreatic hormone-expressing cells in the
zebrafish embryo.";
RL Mech. Dev. 87:217-221(1999).
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL; AJ237750; CAC20109.1; -.
DR HSSP; P01308; 1LPH.
DR ZFIN; ZDB-GENE-980526-110; ins.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005179; F:hormone activity; IEA.
DR GO; GO:0007582; P:physiological processes; IEA.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 53 INSULIN B CHAIN.
FT CHAIN 86 108 INSULIN A CHAIN.
SQ SEQUENCE 108 AA: 11904 MW: 3195289E72AD6D25 CRC64;

```

Query Match          66.5%;  Score 195.5;  DB 13;  Length 108;
Best Local Similarity 45.1%;  Pred. No. 5.1e-19;
Matches 37;  Conservative 5;  Mismatches 7;  Indels 33;  Gaps 1;

Y      4 QHLCGSHLVEALYLVCGERGFFYTPK-----T 30
        |||||||:|||||| | ||| |
D      27 QHLCGSHLVDALYLVCGPTGFYNPKRDVEPLLGFLPPKSAQETEVADFAFKDHAELIRK 86

Y      31 RGIVEQCCTSICSLYQLENYCN 52
        |||||||  ||:::|:|||||
D      87 RGIVEQCCHKPCSIEFLONYCN 108

```

RESULT 6
Q90ZN4
ID Q90ZN4 PRELIMINARY; PRT: 108 AA.

AC Q90ZN4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Preproinsulin.
OS Catla catla (catla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Catla.
OX NCBI_TaxID=72446;
RN [1]
RP SEQUENCE FROM N.A.
RA Bhattacharya S., Roy S.S., Dasgupta S., Ravikumar L., Mukherjee M.,
RA Bandyopadhyaya I., Wakabayasi K.;
RT "A new cell secreting insulin.";
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL; AF373021; AAK51558.1; -.
DR HSSP; P01308; 1LNP.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005179; F:hormone activity; IEA.
DR GO; GO:0007582; P:physiological processes; IEA.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; I1GF; 1.
DR PROSITE; PS00262; INSULIN; 1.
SQ SEQUENCE 108 AA; 11881 MW; D713026E22EF5D59 CRC64;

Query Match 66.5%; Score 195.5; DB 13; Length 108;
Best Local Similarity 45.1%; Pred. No. 5.1e-19;
Matches 37; Conservative 5; Mismatches 7; Indels 33; Gaps 1;

Qy	4 QHLCGSHLVEALYLVCGERGFFYTPK-----	T 30
	:	
Db	27 QHLCGSHLVDALYLVCGPTGFYNPKRDVDPLMGFLPPKSAQETEVADFAFKDHAEVIRK	86
Qy	31 RGIVEQCCTSICSLYQLENYCN 52	
	::: :	
Db	87 RGIVEQCCHKPCSIFELQNYCN 108	

RESULT 7

g98TB0

ID Q98TB0 PRELIMINARY; PRT; 111 AA.

AC Q98TB0;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Preproinsulin (Fragment).

OS Chitala chitala (clown knifefish).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;

OC Osteoglossiformes; Notopteridae; Chitala.

OX NCBI_TaxID=112163;

RN [1] —

RP SEQUENCE FROM N.A.
RX MEDLINE=21203577; PubMed=11306171;
RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
RT osteoglossomorphs and a cyprinid.";
RL Mol. Cell. Endocrinol. 174:51-58(2001).
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL; AF199586; AAK28710.1; -.
DR HSSP; P01308; 1LPH.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005179; F:hormone activity; IEA.
DR GO; GO:0007582; P:physiological processes; IEA.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; ILGF; 1.
FT NON_TER 111 111
SQ SEQUENCE 111 AA; 12483 MW; 247CA4431376329F CRC64;

Query Match 66.3%; Score 195; DB 13; Length 111;
Best Local Similarity 44.2%; Pred. No. 6.1e-19;
Matches 38; Conservative 3; Mismatches 9; Indels 36; Gaps 1;

RESULT 8

Q90ZY1
ID Q90ZY1 PRELIMINARY; PRT; 110 AA.
AC Q90ZY1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Preproinsulin (Fragment).
OS Hiodon alosoides (goldeye).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;
OC Osteoglossiformes; Hiodontidae; Hiodon.
OX NCBI_TaxID=54904;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21203577; PubMed=11306171;
RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
osteoglossomorphs and a cyprinid.";
RL Mol. Cell. Endocrinol. 174:51-58(2001).
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL; AF282408; AAK54684.1; -.
DR HSSP; P01308; 1LNP.

DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR GO; GO:0007582; P:physiological processes; IEA.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; ILGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 FT NON_TER 110 110
 SQ SEQUENCE 110 AA; 12343 MW; BDECCD7703E52E06 CRC64;

 Query Match 65.8%; Score 193.5; DB 13; Length 110;
 Best Local Similarity 42.4%; Pred. No. 9.7e-19;
 Matches 36; Conservative 7; Mismatches 7; Indels 35; Gaps 1;

 Qy 3 NQHLCGSHLVEALYLVCGERGFFYTPKT----- 30
 :|||||||:||:||||:|||| |||
 Db 26 SQHLCGSHLVDALYLYMVCGEKGFYQPKTKRDVDPILLGFLSPKSAQENEADEYPYKDQGDL 85

 Qy 31 ---RGIVEQCCTSICSLYQLENYCN 52
 ||||||| |::: | |||
 Db 86 KVKGIVEQCCHRPCNIFDLNQYCN 110

RESULT 9

Q98TA7
 ID Q98TA7 PRELIMINARY; PRT; 111 AA.
 AC Q98TA7;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Preproinsulin (Fragment).
 OS Osteoglossum bicirrhosum (silver arawana).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;
 OC Osteoglossiformes; Osteoglossidae; Osteoglossum.
 OX NCBI_TaxID=109271;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21203577; PubMed=11306171;
 RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
 RT "Molecular cloning of preproinsulin cDNAs from several
 osteoglossomorphs and a cyprinid.";
 RL Mol. Cell. Endocrinol. 174:51-58(2001).
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 DR EMBL; AF199589; AAK28713.1; -.
 DR HSSP; P01315; 1MPJ.
 DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR GO; GO:0007582; P:physiological processes; IEA.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; ILGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 FT NON_TER 111 111

SQ SEQUENCE 111 AA; 12491 MW; AC9E19D2D4866D20 CRC64;
 Query Match 65.1%; Score 191.5; DB 13; Length 111;
 Best Local Similarity 41.2%; Pred. No. 1.8e-18;
 Matches 35; Conservative 10; Mismatches 5; Indels 35; Gaps 1;

 QY 3 NQHLCGSHLVEALYLVCGERGFFYTPKT----- 30
 ::|||||||:|||:||||:||||:|||:
 Db 27 SQRLCGSHLVDALYMGCGDRGFYSPKSRRAEPLLGFSPKGQENEVDEYPYKEQGEL 86

 QY 31 ---RGIVEQCCTSICSLYQLENYCN 52
 ||||||||| |::: |:|||:
 Db 87 KVKRGIVEQCCHRPCNIFDLQNYCN 111

 RESULT 10
 Q98TA9
 ID Q98TA9 PRELIMINARY; PRT; 87 AA.
 AC Q98TA9;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Preproinsulin (Fragment).
 OS Gnathonemus petersii.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;
 OC Osteoglossiformes; Mormyridae; Gnathonemus.
 OX NCBI_TaxID=42645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21203577; PubMed=11306171;
 RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
 RT "Molecular cloning of preproinsulin cDNAs from several
 osteoglossomorphs and a cyprinid.";
 RL Mol. Cell. Endocrinol. 174:51-58(2001).
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 DR EMBL; AF199587; AAK28711.1; -.
 DR HSSP; P01308; 1HIS.
 DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR GO; GO:0007582; P:physiological processes; IEA.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; ILGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 FT NON_TER 1 1
 FT NON_TER 87 87
 SQ SEQUENCE 87 AA; 9874 MW; FF448ED35D2453F5 CRC64;

 Query Match 63.8%; Score 187.5; DB 13; Length 87;
 Best Local Similarity 42.9%; Pred. No. 5.1e-18;
 Matches 36; Conservative 5; Mismatches 8; Indels 35; Gaps 1;

 QY 4 QHLCGSHLVEALYLVCGERGFFYTPKT----- 30
 |||||||||:|||||:|||: | |

```

Db        4 QHLCGSHLVEALFLVCGERGFFNPDTKRDVDSLGLSPKSGPENEADEYRYKEQAEVK 63
Qy        31 --RGIVEQCCTSICSLYQLENYCN 52
           ||||||| |::: | |||
Db        64 VKRGIVEQCCHHPCNIFDLNQYCN 87

```

RESULT 11

Q98TB1

```

ID  Q98TB1      PRELIMINARY;      PRT;    108 AA.
AC  Q98TB1;
DT  01-JUN-2001 (TrEMBLrel. 17, Created)
DT  01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT  01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE  Preproinsulin (Fragment).
OS  Catostomus commersoni (White sucker).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC  Catostomidae; Catostomus.
OX  NCBI_TaxID=7971;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=21203577; PubMed=11306171;
RA  Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H. ;
RT  "Molecular cloning of preproinsulin cDNAs from several
RT  osteoglossomorphs and a cyprinid." ;
RL  Mol. Cell. Endocrinol. 174:51-58(2001).
CC  -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC  -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR  EMBL; AF199585; AAK28709.1; -.
DR  HSSP; P01308; 1LPH.
DR  GO; GO:0005576; C:extracellular; IEA.
DR  GO; GO:0005179; F:hormone activity; IEA.
DR  GO; GO:0007582; P:physiological processes; IEA.
DR  InterPro; IPR004825; Ins/IGF/relax.
DR  Pfam; PF00049; Insulin; 1.
DR  PRINTS; PR00277; INSULINB.
DR  SMART; SM00078; ILGF; 1.
DR  PROSITE; PS00262; INSULIN; 1.
FT  NON_TER      108      108
SQ  SEQUENCE      108 AA;  11873 MW;  E426310696FBAFC8 CRC64;
```

```

Query Match          63.4%;  Score 186.5;  DB 13;  Length 108;
Best Local Similarity 43.9%;  Pred. No. 8.7e-18;
Matches   36;  Conservative  4;  Mismatches  9;  Indels   33;  Gaps     1;

```

```

Qy        4 QHLCGSHLVEALYLVCGERGFFYTPK-----T 30
           |||||||:|||||| | |||| |
Db        27 QHLCGSHLVDALYLVCGPTGFFYNPKRDVDPLIGFLPPKSGPENEADFAFKDHAEIRK 86
Qy        31 RGIVEQCCTSICSLYQLENYCN 52
           ||||||| |::: || |||
Db        87 RGIVEQCCHRPNIFDLEKYCN 108

```

RESULT 12

Q98TB2

ID Q98TB2 PRELIMINARY; PRT; 91 AA.
 AC Q98TB2;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Preproinsulin (Fragment).
 OS Ambloplites rupestris (Rock bass).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidei;
 OC Centrarchidae; Ambloplites.
 OX NCBI_TaxID=109273;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Al-Mahrouki A.A., Irwin D.M., Youson J.H.;
 RT "Molecular cloning of preproinsulin cDNA from the rock bass.";
 RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 DR EMBL; AF199584; AAK28708.1; -.
 DR HSSP; P01308; 1LPH.
 DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR GO; GO:0007582; P:physiological processes; IEA.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; ILGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 FT NON_TER 1 1
 FT NON_TER 91 91
 SQ SEQUENCE 91 AA; 10100 MW; E86C8B256DC69D39 CRC64;

Query Match 63.1%; Score 185.5; DB 13; Length 91;
 Best Local Similarity 40.9%; Pred. No. 1e-17;
 Matches 36; Conservative 5; Mismatches 8; Indels 39; Gaps 1;

Qy	4 QHLCGSHLVEALYLVCGERGFFYTPK-----	29
	: :	
Db	4 QHLCGSHLVDALYLVCGDRGFFYNPKRDVDPIMGFLPPKADGAAAPGNEVAEFAFKDQ	63
Qy	30 -----TRGIVEQCCTSICSLYQLENYCN 52	
	:::	
Db	64 MEMMVKRGIVEQCCHHPCNIFDLGRYCN 91	

RESULT 13

Q8HZ81

ID Q8HZ81 PRELIMINARY; PRT; 65 AA.
 AC Q8HZ81;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Insulin (Fragment).
 OS Gorilla gorilla (gorilla).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.

OX NCBI_TaxID=9593;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA O'hUigin C., Tichy H., Klein J.;
 RT "Molecular evolution in higher primates; gene specific and organism
 RT specific characteristics.";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AY092023; AAM76640.1; -.
 DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR GO; GO:0007582; P:physiological processes; IEA.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR SMART; SM00078; IlGF; 1.
 FT NON_TER 1 1
 FT NON_TER 65 65
 SQ SEQUENCE 65 AA; 6920 MW; B772017FD8BCABEA CRC64;

Query Match 49.7%; Score 146; DB 6; Length 65;
 Best Local Similarity 47.7%; Pred. No. 1.9e-12;
 Matches 31; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

Qy	7 CGSHLVEALYLVCGERGFFYTPKT-----	RG 32
Db	1 CGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEGSLQKRG 60	
Qy	33 IVEQC 37	
Db	61 IVEQC 65	

RESULT 14

Q8HZ80
ID Q8HZ80 PRELIMINARY; PRT; 65 AA.
AC Q8HZ80;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Insulin (Fragment).
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX NCBI_TaxID=9600;
RN [1]
RP SEQUENCE FROM N.A.
RA O'hUigin C., Tichy H., Klein J.;
RT "Molecular evolution in higher primates; gene specific and organism
specific characteristics.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY092024; AAM76641.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005179; F:hormone activity; IEA.
DR GO; GO:0007582; P:physiological processes; IEA.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR SMART; SM00078; ILGF; 1.
FT NON_TER 1 1

FT NON_TER 65 65
SQ SEQUENCE 65 AA; 6920 MW; B772017FD8BCABEA CRC64;

Query Match 49.7%; Score 146; DB 6; Length 65;
Best Local Similarity 47.7%; Pred. No. 1.9e-12;
Matches 31; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

Qy 7 CGSHLVEALYLVCGERGFFYTPKT-----RG 32
Db 1 CGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEGSLOKRG 60

Qy 33 IVEQC 37
Db 61 IVEQC 65

RESULT 15

Q90XD0

ID Q90XD0 PRELIMINARY; PRT; 207 AA.
AC Q90XD0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Preproinsulin-like growth factor-II.
GN IGF-II.
OS Cyprinus carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Cyprinus.
OX NCBI_TaxID=7962;
RN [1]
RP SEQUENCE FROM N.A.
RA Tse M.C.L., Chan K.M., Cheng C.H.K.;
RT "PCR-cloning and Gene Expression Studies on Common Carp (Cyprinus
carpio) Insulin-like Growth Factor-II.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL; AF402958; AAL25799.1; -.
DR HSSP; P01308; 1LNP.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005179; F:hormone activity; IEA.
DR GO; GO:0007582; P:physiological processes; IEA.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
SQ SEQUENCE 207 AA; 23869 MW; 44FBA2871A361862 CRC64;

Query Match 49.0%; Score 144; DB 13; Length 207;
Best Local Similarity 48.2%; Pred. No. 1.2e-11;
Matches 27; Conservative 8; Mismatches 11; Indels 10; Gaps 1;

Qy 6 LCGSHLVEALYLVCGERGFFYTPKT-----RGIVEQCCTSICSLYQLENYC 51
Db 53 LCGGELVDALQFVCGDRGFYFSRPTSRLSSRRSQRNGIVEECCFNSCNLALLEQYC 108

Search completed: July 15, 2004, 16:41:00
Job time : 21.9204 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2004, 16:28:49 ; Search time 3.58955 Seconds
(without alignments)
754.314 Million cell updates/sec

Title: US-09-423-100-5

Perfect score: 294

Sequence: 1 FVNQHLCGSHLVEALYLVCG.....IVEQCCTSICSLYQLENYCN 52

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

%

Result	Query				Description
No.	Score	Match	Length	DB ID	
1	273.5	93.0	51	1 INS_BALPH	P01312 balaenopter
2	273.5	93.0	51	1 INS_ELEMA	P01316 elephas max
3	268.5	91.3	51	1 INS_ACOCA	P01324 acomys cahi
4	267	90.8	110	1 INS_CERAE	P30407 cercopithec
5	267	90.8	110	1 INS_HUMAN	P01308 homo sapien
6	267	90.8	110	1 INS_MACFA	P30406 macaca fasc
7	267	90.8	110	1 INS_PANTR	P30410 pan troglod
8	263.5	89.6	51	1 INS_BALBO	P01314 balaenopter
9	263.5	89.6	51	1 INS_CAMDR	P01320 camelus dro
10	263.5	89.6	51	1 INS_CAPII	P01319 capra hircu
11	263	89.5	108	1 INS_PIG	P01315 sus scrofa
12	263	89.5	110	1 INS_RABIT	P01311 oryctolagus
13	263	89.5	110	1 INS_SPETR	Q91xi3 spermophilu
14	262.5	89.3	51	1 INS_FELCA	P06306 felis silve
15	262	89.1	110	1 INS_CANFA	P01321 canis famil
16	260	88.4	110	1 INS_CRILO	P01313 cricetus
17	258.5	87.9	105	1 INS_BOVIN	P01317 bos taurus

18	257	87.4	110	1	INS_PSAOB	Q62587 psammomys o
19	256.5	87.2	51	1	INS_DIDMA	P18109 didelphis m
20	254.5	86.6	105	1	INS_SHEEP	P01318 ovis aries
21	252	85.7	86	1	INS_HORSE	P01310 equus cabal
22	251.5	85.5	51	1	INS_CHIBR	P01327 chinchilla
23	251	85.4	108	1	INS_AOTTR	P10604 aotus trivi
24	250	85.0	108	1	INS1_MOUSE	P01325 mus musculu
25	249	84.7	110	1	INS1_RAT	P01322 rattus norv
26	248.5	84.5	51	1	INS_ANSAN	P07454 anser anser
27	248	84.4	110	1	INS2_MOUSE	P01326 mus musculu
28	248	84.4	110	1	INS2_RAT	P01323 rattus norv
29	246	83.7	52	1	INS_ACIGU	P81423 acipenser g
30	244.5	83.2	51	1	INS_HYSCR	P01328 hystrix cri
31	244.5	83.2	51	1	INS_TRASC	P31887 trachemys s
32	238.5	81.1	103	1	INS_SELRF	P51463 selaphorus
33	235.5	80.1	51	1	INS_ORNAN	Q9tqy7 ornithorhyn
34	234.5	79.8	107	1	INS_CHICK	P01332 gallus gall
35	233.5	79.4	81	1	INS_ANAPL	P01333 anas platyr
36	231.5	78.7	51	1	INS_ALLMI	P12703 alligator m
37	231	78.6	52	1	INS_LEPSP	P09476 lepisosteus
38	228.5	77.7	51	1	INS_ZAODH	P12708 zaocys dhum
39	227.5	77.4	51	1	INS_CROAT	P01334 crotalus at
40	226.5	77.0	106	1	INS1_XENLA	P12706 xenopus lae
41	226.5	77.0	106	1	INS2_XENLA	P12707 xenopus lae
42	222	75.5	57	1	INS_PETMA	P14806 petromyzon
43	216	73.5	51	1	INS_PLAFE	P09477 platichthys
44	214	72.8	52	1	INS_AMICA	P29335 amia calva
45	211	71.8	50	1	INS_ONCGO	P23187 oncorhynchus

ALIGNMENTS

RESULT 1
 INS_BALPH
 ID INS_BALPH STANDARD; PRT; 51 AA.
 AC P01312;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Insulin.
 GN INS.
 OS Balaenoptera physalus (Finback whale) (Common rorqual), and
 OS Physeter catodon (Sperm whale) (Physeter macrocephalus).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
 OC Balaenopteridae; Balaenoptera.
 OX NCBI_TaxID=9770, 9755;
 RN [1]
 RP PARTIAL SEQUENCE.
 RC SPECIES=B.physalus;
 RA Hama H., Titani K., Sakaki S., Narita K.;
 RT "The amino acid sequence in fin-whale insulin.";
 RL J. Biochem. 56:285-293(1964).
 RN [2]
 RP SEQUENCE.
 RC SPECIES=P.catodon;

RA Ishihara Y., Saito T., Ito Y., Fujino M.;
 RT "Structure of sperm- and sei-whale insulins and their breakdown by
 whale pepsin.";
 RL Nature 181:1468-1469(1958).
 RN [3]
 RP SEQUENCE.
 RC SPECIES=P.catodon;
 RA Harris J.I., Sanger F., Naughton M.A.;
 RT "Species differences in insulin.";
 RL Arch. Biochem. Biophys. 65:427-438(1956).
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
 increases cell permeability to monosaccharides, amino acids and
 fatty acids. It accelerates glycolysis, the pentose phosphate
 cycle, and glycogen synthesis in liver.
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
 disulfide bonds.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to the insulin family.
 DR PIR; A91918; INWHF.
 DR PIR; A93142; INWHP.
 DR HSSP; P01317; 1APH.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; IIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism.
 FT CHAIN 1 30 INSULIN B CHAIN.
 FT NON_CONS 30 31
 FT CHAIN 31 51 INSULIN A CHAIN.
 FT DISULFID 7 37 INTERCHAIN.
 FT DISULFID 19 50 INTERCHAIN.
 FT DISULFID 36 41
 SQ SEQUENCE 51 AA; 5766 MW; 9007B514691A7CDD CRC64;

 Query Match 93.0%; Score 273.5; DB 1; Length 51;
 Best Local Similarity 96.2%; Pred. No. 1.9e-27;
 Matches 50; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

 Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
 ||||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCTSICSLYQLENYCN 51

RESULT 2
 INS_ELEMA
 ID INS_ELEMA STANDARD; PRT; 51 AA.
 AC P01316;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Insulin.
 GN INS.
 OS Elephas maximus (Indian elephant).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Proboscidea; Elephantidae; Elephas.
 OX NCBI_TaxID=9783;
 RN [1]

RP SEQUENCE.
 RX MEDLINE=66160119; PubMed=5949593;
 RA Smith L.F.;
 RT "Species variation in the amino acid sequence of insulin.";
 RL Am. J. Med. 40:662-666(1966).
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
 increases cell permeability to monosaccharides, amino acids and
 fatty acids. It accelerates glycolysis, the pentose phosphate
 cycle, and glycogen synthesis in liver.
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
 disulfide bonds.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- MISCELLANEOUS: THE SPECIES OF ELEPHANT IS NOT GIVEN, BUT IT IS
 MOST PROBABLY THE INDIAN ELEPHANT (ELEPHAS MAXIMUS).
 CC -!- SIMILARITY: Belongs to the insulin family.
 DR HSSP; P01308; 1AI0.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; ILGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism.
 FT CHAIN 1 30 INSULIN B CHAIN.
 FT NON_CONS 30 31
 FT CHAIN 31 51 INSULIN A CHAIN.
 FT DISULFID 7 37 INTERCHAIN.
 FT DISULFID 19 50 INTERCHAIN.
 FT DISULFID 36 41
 SQ SEQUENCE 51 AA; 5752 MW; 9007B50CDB457D6D CRC64;

 Query Match 93.0%; Score 273.5; DB 1; Length 51;
 Best Local Similarity 94.2%; Pred. No. 1.9e-27;
 Matches 49; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

 Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
 ||||||| ||||| ||||| ||||| ||||| :||||| |||||
 Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT-GIVEQCCTGVCSLYQLENYCN 51

RESULT 3
 INS_ACOCA
 ID INS_ACOCA STANDARD; PRT; 51 AA.
 AC P01324;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Insulin.
 GN INS.
 OS Acomys cahirinus (Egyptian spiny mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Acomys.
 OX NCBI_TaxID=10068;
 RN [1]
 RP COMPOSITION.
 RX MEDLINE=72189454; PubMed=5028210;
 RA Buerzli H.F., Humbel R.E.;
 RT "Isolation and partial structural analysis of insulin from mouse (Mus
 musculus) and spiny mouse (Acomys cahirinus).";

RL Hoppe-Seyler's Z. Physiol. Chem. 353:444-450(1972).
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
 CC increases cell permeability to monosaccharides, amino acids and
 CC fatty acids. It accelerates glycolysis, the pentose phosphate
 CC cycle, and glycogen synthesis in liver.
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
 CC disulfide bonds.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to the insulin family.
 DR PIR; A01591; INMSSP.
 DR HSSP; P01308; 1TYM.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; ILGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism.
 FT CHAIN 1 30 INSULIN B CHAIN.
 FT NON_CONS 30 31
 FT CHAIN 31 51 INSULIN A CHAIN.
 FT DISULFID 7 37 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 19 50 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 36 41 BY SIMILARITY.
 SQ SEQUENCE 51 AA; 5768 MW; 992BD8B629047D3D CRC64;
 Query Match 91.3%; Score 268.5; DB 1; Length 51;
 Best Local Similarity 92.3%; Pred. No. 7.8e-27;
 Matches 48; Conservative 3; Mismatches 0; Indels 1; Gaps 1;
 Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
 ||:|||||:|||||:|||||:|||||:|||||:|||||:
 Db 1 FVBQHLCGSHLVEALYLVCGERGFFYTPKS-GIVDQCCTSICSLYQLENYCN 51

RESULT 4
 INS_CERAE
 ID INS_CERAE STANDARD; PRT; 110 AA.
 AC P30407; P01309;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Insulin precursor.
 GN INS.
 OS Cercopithecus aethiops (Green monkey) (Grivet).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
 OC Cercopithecinae; Cercopithecus.
 OX NCBI_TaxID=9534;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92219953; PubMed=1560757;
 RA Seino S., Bell G.I., Li W.;
 RT "Sequences of primate insulin genes support the hypothesis of a
 RT slower rate of molecular evolution in humans and apes than in
 RT monkeys.";
 RL Mol. Biol. Evol. 9:193-203(1992).
 RN [2]
 RP SEQUENCE OF 57-87.

RX MEDLINE=72258016; PubMed=4626369;
 RA Peterson J.D., Nehrlich S., Oyer P.E., Steiner D.F.;
 RT "Determination of the amino acid sequence of the monkey, sheep, and
 RT dog proinsulin C-peptides by a semi-micro Edman degradation
 RT procedure.";
 RL J. Biol. Chem. 247:4866-4871(1972).
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
 CC increases cell permeability to monosaccharides, amino acids and
 CC fatty acids. It accelerates glycolysis, the pentose phosphate
 CC cycle, and glycogen synthesis in liver.
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
 CC disulfide bonds.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to the insulin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X61092; CAA43405.1; -.
 DR PIR; B42179; B42179.
 DR HSSP; P01308; 1AI0.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; IIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 54 INSULIN B CHAIN.
 FT PROPEP 57 87 C PEPTIDE.
 FT CHAIN 90 110 INSULIN A CHAIN.
 FT DISULFID 31 96 INTERCHAIN.
 FT DISULFID 43 109 INTERCHAIN.
 FT DISULFID 95 100
 SQ SEQUENCE 110 AA; 12019 MW; 95A1F54BE7B247F9 CRC64;
 Query Match 90.8%; Score 267; DB 1; Length 110;
 Best Local Similarity 60.5%; Pred. No. 2.6e-26;
 Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;
 Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 30
 |||||||
 Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGPGAGSLQPLALEG 84
 Qy 31 ----RGIVEQCCTSICSLYQLENYCN 52
 |||||||
 Db 85 SLQKRGIVEQCCTSICSLYQLENYCN 110
 RESULT 5
 INS_HUMAN
 ID INS_HUMAN STANDARD; PRT; 110 AA.

AC P01308;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Insulin precursor.
GN INS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=80120725; PubMed=6243748;
RA Bell G.I., Pictet R.L., Rutter W.J., Cordell B., Tischer E.,
RA Goodman H.M.;
RT "Sequence of the human insulin gene.";
RL Nature 284:26-32(1980).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=80236313; PubMed=6248962;
RA Ullrich A., Dull T.J., Gray A., Brosius J., Sures I.;
RT "Genetic variation in the human insulin gene.";
RL Science 209:612-615(1980).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=80054779; PubMed=503234;
RA Bell G.I., Swain W.F., Pictet R.L., Cordell B., Goodman H.M.,
RA Rutter W.J.;
RT "Nucleotide sequence of a cDNA clone encoding human preproinsulin.";
RL Nature 282:525-527(1979).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=80147417; PubMed=6927840;
RA Sures I., Goeddel D.V., Gray A., Ullrich A.;
RT "Nucleotide sequence of human preproinsulin complementary DNA.";
RL Science 208:57-59(1980).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=93364428; PubMed=8358440;
RA Lucassen A.M., Bell J.I., Julier C., Lathrop M.;
RT "Susceptibility to insulin dependent diabetes mellitus maps to a 4.1
RT kb segment of DNA spanning the insulin gene and associated VNTR.";
RL Nat. Genet. 4:305-310(1993).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnurch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [7]
RP SEQUENCE OF 1-59 FROM N.A.
RC TISSUE=Blood;
RA Fajardy I.I., Weill J.J., Stuckens C.C., Danze P.M.P.;
RT "Description of a novel RFLP diallelic polymorphism (-127 BsgI C/G)
RT within the 5' region of insulin gene.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN [8]
RP SEQUENCE OF 25-54 AND 90-110.
RA Nicol D.S.H.W., Smith L.F.;
RT "Amino-acid sequence of human insulin.";
RL Nature 187:483-485(1960).
RN [9]
RP SEQUENCE OF 57-87.
RX MEDLINE=71116410; PubMed=5101771;
RA Oyer P.E., Cho S., Peterson J.D., Steiner D.F.;
RT "Studies on human proinsulin. Isolation and amino acid sequence of
RT the human pancreatic C-peptide.";
RL J. Biol. Chem. 246:1375-1386(1971).
RN [10]
RP SEQUENCE OF 57-87.
RX MEDLINE=71257722; PubMed=5560404;
RA Ko A., Smyth D.G., Markussen J., Sundby F.;
RT "The amino acid sequence of the C-peptide of human proinsulin.";
RL Eur. J. Biochem. 20:190-199(1971).
RN [11]
RP SYNTHESIS.
RX MEDLINE=75077277; PubMed=4443293;
RA Sieber P., Kamber B., Hartmann A., Joehl A., Riniker B., Rittel W.;
RT "Total synthesis of human insulin under directed formation of the
RT disulfide bonds.";
RL Helv. Chim. Acta 57:2617-2621(1974).
RN [12]
RP SYNTHESIS OF 57-87.
RX MEDLINE=75040007; PubMed=4803504;
RA Naithani V.K.;
RT "Studies on polypeptides, IV. The synthesis of C-peptide of human
RT proinsulin.";
RL Hoppe-Seyler's Z. Physiol. Chem. 354:659-672(1973).
RN [13]
RP SYNTHESIS OF 65-69 AND 70-73.
RX MEDLINE=73161263; PubMed=4698555;
RA Geiger R., Volk A.;
RT "Synthesis of peptides with the properties of human proinsulin C
RT peptides (hC peptide). 3. Synthesis of the sequences 14-17 and 9-13
RT of human proinsulin C peptides.";
RL Chem. Ber. 106:199-205(1973).

RN [14]
RP SYNTHESIS OF 84-87.
RX MEDLINE=73161261; PubMed=4698553;
RA Geiger R., Jaeger G., Keonig W., Treuth G.;
RT "Synthesis of peptides with the properties of human proinsulin C
peptides (hC peptide). I. Scheme for the synthesis and preparation of
the sequence 28-31 of human proinsulin C peptide.";
RL Chem. Ber. 106:188-192(1973).
RN [15]
RP VARIANT LOS ANGELES SER-48.
RX MEDLINE=84016053; PubMed=6312455;
RA Haneda M., Chan S.J., Kwok S.C.M., Rubenstein A.H., Steiner D.F.;
RT "Studies on mutant human insulin genes: identification and sequence
analysis of a gene encoding [SerB24]insulin.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:6366-6370(1983).
RN [16]
RP VARIANTS LOS ANGELES SER-48 AND CHICAGO LEU-49.
RX MEDLINE=84170233; PubMed=6424111;
RA Shoelson S., Fickova M., Haneda M., Nahum A., Musso G., Kaiser E.T.,
RA Rubenstein A.H., Tager H.;
RT "Identification of a mutant human insulin predicted to contain a
RT serine-for-phenylalanine substitution.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:7390-7394(1983).
RN [17]
RP VARIANT PROVIDENCE ASP-34.
RX MEDLINE=87175640; PubMed=3470784;
RA Chan S.J., Seino S., Gruppuso P.A., Schwartz R., Steiner D.F.;
RT "A mutation in the B chain coding region is associated with impaired
RT proinsulin conversion in a family with hyperproinsulinemia.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:2194-2197(1987).
RN [18]
RP VARIANT WAKAYAMA LEU-92.
RX MEDLINE=87058122; PubMed=3537011;
RA Sakura H., Iwamoto Y., Sakamoto Y., Kuzuya T., Hirata H.;
RT "Structurally abnormal insulin in a diabetic patient. Characterization
RT of the mutant insulin A3 (Val-->Leu) isolated from the pancreas.";
RL J. Clin. Invest. 78:1666-1672(1986).
RN [19]
RP VARIANT HIS-89.
RX MEDLINE=90317021; PubMed=2196279;
RA Barbetti F., Raben N., Kadowaki T., Cama A., Accili D., Gabbay K.H.,
RA Merenich J.A., Taylor S.I., Roth J.;
RT "Two unrelated patients with familial hyperproinsulinemia due to a
RT mutation substituting histidine for arginine at position 65 in the
RT proinsulin molecule: identification of the mutation by direct
RT sequencing of genomic deoxyribonucleic acid amplified by polymerase
RT chain reaction.";
RL J. Clin. Endocrinol. Metab. 71:164-169(1990).
RN [20]
RP VARIANT HIS-89.
RX MEDLINE=85261996; PubMed=4019786;
RA Shibasaki Y., Kawakami T., Kanazawa Y., Akanuma Y., Takaku F.;
RT "Posttranslational cleavage of proinsulin is blocked by a point
RT mutation in familial hyperproinsulinemia.";
RL J. Clin. Invest. 76:378-380(1985).
RN [21]
RP VARIANT KYOTO LEU-89.

RX MEDLINE=92291307; PubMed=1601997;
RA Yano H., Kitano N., Morimoto M., Polonsky K.S., Imura H., Seino Y.;
RT "A novel point mutation in the human insulin gene giving rise to
RT hyperproinsulinemia (proinsulin Kyoto).";
RL J. Clin. Invest. 89:1902-1907(1992).
RN [22]
RP STRUCTURE BY NMR.
RX MEDLINE=91104966; PubMed=2271664;
RA Hua Q.-X., Weiss M.A.;
RT "Toward the solution structure of human insulin: sequential 2D 1H NMR
RT assignment of a des-pentapeptide analogue and comparison with crystal
RT structure.";
RL Biochemistry 29:10545-10555(1990).
RN [23]
RP STRUCTURE BY NMR.
RX MEDLINE=91242467; PubMed=2036420;
RA Hua Q.-X., Weiss M.A.;
RT "Comparative 2D NMR studies of human insulin and des-pentapeptide
RT insulin: sequential resonance assignment and implications for protein
RT dynamics and receptor recognition.";
RL Biochemistry 30:5505-5515(1991).
RN [24]
RP STRUCTURE BY NMR.
RX MEDLINE=91265527; PubMed=1646635;
RA Hua Q.-X., Weiss M.A.;
RT "Two-dimensional NMR studies of Des-(B26-B30)-insulin: sequence-
RT specific resonance assignments and effects of solvent composition.";
RL Biochim. Biophys. Acta 1078:101-110(1991).

Query Match 90.8%; Score 267; DB 1; Length 110;
Best Local Similarity 60.5%; Pred. No. 2.6e-26;
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

Qy	1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT-----	30
Db	25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG	84
Qy	31 ----RGIVEQCCTSICSLYQLENYCN 52	
Db	85 SLQKRGIVEQCCTSICSLYQLENYCN 110	

RESULT 6

INS MACFA

ID - INS MACFA STANDARD; PRT; 110 AA.

AC P30406; P01309;

DT 21-JUL-1986 (Rel. 01, Created)

DT 13-AUG-1987 (Rel. 05, Last sequence update)

DT 10-OCT-2003 (Rel.

DE Insu.

GN INS.
OS *Macaca fascicularis* (Crab eating macaque) (*Cynomolgus* monkey).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi

OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;

OC Cercopithecinæ; Macaca.

OX NCBT TaxID=9541;

RN [1] =

11

RP SEQUENCE FROM N.A.
 RX MEDLINE=83080474; PubMed=6184262;
 RA Wetekam W., Groneberg J., Leineweber M., Wengenmayer F.,
 RA Winnacker E.-L.;
 RT "The nucleotide sequence of cDNA coding for preproinsulin from the
 RT primate Macaca fascicularis.";
 RL Gene 19:179-183(1982).
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
 CC increases cell permeability to monosaccharides, amino acids and
 CC fatty acids. It accelerates glycolysis, the pentose phosphate
 CC cycle, and glycogen synthesis in liver.
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
 CC disulfide bonds.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to the insulin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; J00336; AAA36849.1; -.
 DR PIR; JQ0178; JQ0178.
 DR HSSP; P01308; 1AI0.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; I1GF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 54 INSULIN B CHAIN.
 FT PROPEP 57 87 C PEPTIDE.
 FT CHAIN 90 110 INSULIN A CHAIN.
 FT DISULFID 31 96 INTERCHAIN.
 FT DISULFID 43 109 INTERCHAIN.
 FT DISULFID 95 100
 SQ SEQUENCE 110 AA; 11991 MW; 83C6E33A80A420F9 CRC64;

 Query Match 90.8%; Score 267; DB 1; Length 110;
 Best Local Similarity 60.5%; Pred. No. 2.6e-26;
 Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

 Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 30
 Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGPGAGSLQPLALEG 84

 Qy 31 ----RGIVEQCCTSICSLYQLENYCN 52
 Db 85 SLQKRGIVEQCCTSICSLYQLENYCN 110

RESULT 7
 INS_PANTR

ID INS_PANTR STANDARD; PRT; 110 AA.

AC P30410;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Insulin precursor.

GN INS.

OS Pan troglodytes (Chimpanzee).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

OX NCBI_TaxID=9598;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92219953; PubMed=1560757;

RA Seino S., Bell G.I., Li W.;

RT "Sequences of primate insulin genes support the hypothesis of a slower rate of molecular evolution in humans and apes than in monkeys.";

RL Mol. Biol. Evol. 9:193-203(1992).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=22833521; PubMed=12952878;

RA Stead J.D., Hurles M.E., Jeffreys A.J.;

RT "Global haplotype diversity in the human insulin gene region.";

RL Genome Res. 13:2101-2111(2003).

CC -!- FUNCTION: Insulin decreases blood glucose concentration. It increases cell permeability to monosaccharides, amino acids and fatty acids. It accelerates glycolysis, the pentose phosphate cycle, and glycogen synthesis in liver.

CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two disulfide bonds.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- SIMILARITY: Belongs to the insulin family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X61089; CAA43403.1; -.

DR EMBL; AY137497; AAN06933.1; -.

DR PIR; A42179; A42179.

DR PDB; 1EFE; 29-MAR-00.

DR InterPro; IPR004825; Ins/IGF/relax.

DR Pfam; PF00049; Insulin; 1.

DR PRINTS; PR00277; INSULINB.

DR SMART; SM00078; ILGF; 1.

DR PROSITE; PS00262; INSULIN; 1.

KW Insulin family; Hormone; Glucose metabolism; Signal; 3D-structure.

FT SIGNAL 1 24

FT CHAIN 25 54 INSULIN B CHAIN.

FT PROPEP 57 87 C PEPTIDE.

FT CHAIN 90 110 INSULIN A CHAIN.

FT DISULFID 31 96 INTERCHAIN.

FT DISULFID 43 109 INTERCHAIN.
 FT DISULFID 95 100
 SQ SEQUENCE 110 AA; 12025 MW; 41EB8DF79837CEF5 CRC64;
 Query Match 90.8%; Score 267; DB 1; Length 110;
 Best Local Similarity 60.5%; Pred. No. 2.6e-26;
 Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;
 QY 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTE----- 30
 |||||||
 Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGPGAGSLQPLALEG 84
 QY 31 ----RGIVEQCCTSICSLYQLENYCN 52
 |||||||
 Db 85 SLQKRGIVEQCCTSICSLYQLENYCN 110

RESULT 8
 INS_BALBO
 ID INS_BALBO STANDARD; PRT; 51 AA.
 AC P01314;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Insulin.
 GN INS.
 OS Balaenoptera borealis (Sei whale).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
 OC Balaenopteridae; Balaenoptera.
 OX NCBI_TaxID=9768;
 RN [1]
 RP SEQUENCE.
 RA Ishihara Y., Saito T., Ito Y., Fujino M.;
 RT "Structure of sperm- and sei-whale insulins and their breakdown by
 whale pepsin.";
 RL Nature 181:1468-1469(1958).
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
 increases cell permeability to monosaccharides, amino acids and
 fatty acids. It accelerates glycolysis, the pentose phosphate
 cycle, and glycogen synthesis in liver.
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
 disulfide bonds.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to the insulin family.
 DR PIR; A01582; INWH1S.
 DR HSSP; P01317; 1APH.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; IIIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism.
 FT CHAIN 1 30 INSULIN B CHAIN.
 FT NON_CONS 30 31
 FT CHAIN 31 51 INSULIN A CHAIN.
 FT DISULFID 7 37 INTERCHAIN.
 FT DISULFID 19 50 INTERCHAIN.

FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5723 MW; 9007B50E400A7DDD CRC64;
Query Match 89.6%; Score 263.5; DB 1; Length 51;
Best Local Similarity 92.3%; Pred. No. 3.3e-26;
Matches 48; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
| ||||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCASTCSILYQLENYCN 51

RESULT 9
INS_CAMDR
ID INS_CAMDR STANDARD; PRT; 51 AA.
AC P01320;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Insulin.
GN INS.
OS Camelus dromedarius (Dromedary) (Arabian camel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
OX NCBI_TaxID=9838;
RN [1]
RP SEQUENCE.
RA Danho W.O.;
RT "The isolation and characterization of insulin of camel (Camelus
dromedarius).";
RL J. Fac. Med. Baghdad 14:16-28(1972).
CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
increases cell permeability to monosaccharides, amino acids and
fatty acids. It accelerates glycolysis, the pentose phosphate
cycle, and glycogen synthesis in liver.
CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
disulfide bonds.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the insulin family.
DR PIR; A92782; INCMA.
DR HSSP; P01317; 2INS.
DR InterPro; IPR004825; Ins/IGF/relax.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5693 MW; 901E88BA085A7DDD CRC64;
Query Match 89.6%; Score 263.5; DB 1; Length 51;
Best Local Similarity 90.4%; Pred. No. 3.3e-26;
Matches 47; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
| :||||||||||||||||| :|||||| :|||||||||
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCASVCSLYQLENYCN 51

RESULT 10
INS_CAPII
ID INS_CAPII STANDARD; PRT; 51 AA.
AC P01319;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Insulin.
GN INS.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE.
RX MEDLINE=66160119; PubMed=5949593;
RA Smith L.F.;
RT "Species variation in the amino acid sequence of insulin."
RL Am. J. Med. 40:662-666(1966).
CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
increases cell permeability to monosaccharides, amino acids and
fatty acids. It accelerates glycolysis, the pentose phosphate
cycle, and glycogen synthesis in liver.
CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
CC disulfide bonds.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the insulin family.
DR PIR; A01586; INGT.
DR HSSP; P01317; 1APH.
DR InterPro; IPR004825; Ins/IGF/relax.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5692 MW; 9007B50CDB4E7DDD CRC64;

Query Match 89.6%; Score 263.5; DB 1; Length 51;
Best Local Similarity 90.4%; Pred. No. 3.3e-26;
Matches 47; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52
| :||||||||||||||||| :|||||| :|||||||||
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCAGVCSLYQLENYCN 51

RESULT 11

INS_PIG

ID INS_PIG STANDARD; PRT; 108 AA.

AC P01315; Q9TSJ5;

DT 21-JUL-1986 (Rel. 01, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Insulin precursor.

GN INS.

OS Sus scrofa (Pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

OX NCBI_TaxID=9823;

RN [1]

RP SEQUENCE FROM N.A.

RA Han X.G., Tuch B.E.;

RT "Complete porcine preproinsulin cDNA sequence.";

RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Large white;

RX MEDLINE=22135958; PubMed=12140686;

RA Amarger V., Nguyen M., Laere A.S., Braunschweig M., Nezer C.,

RA Georges M., Andersson L.;

RT "Comparative sequence analysis of the INS-IGF2-H19 gene cluster in

RT pigs.";

RL Mamm. Genome 13:388-398(2002).

RN [3]

RP SEQUENCE OF 25-108.

RX MEDLINE=68286485; PubMed=5657063;

RA Chance R.E., Ellis R.M., Bromer W.W.;

RT "Porcine proinsulin: characterization and amino acid sequence.";

RL Science 161:165-167(1968).

RN [4]

RP REVISION TO 59.

RA Chance R.E.;

RL Submitted (JUL-1970) to the PIR data bank.

RN [5]

RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).

RA Blundell T.L., Dodson G.G., Hodgkin D., Mercola D.;

RT "Insulin. The structure in the crystal and its reflection in

RT chemistry and biology.";

RL Adv. Protein Chem. 26:279-402(1972).

RN [6]

RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).

RA Isaacs N.W., Agarwal R.C.;

RT "Experience with fast Fourier least squares in the refinement of the

RT crystal structure of rhombohedral 2-zinc insulin at 1.5-A

RT resolution.";

RL Acta Crystallogr. A 34:782-791(1978).

RN [7]

RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).

RX MEDLINE=89099318; PubMed=2905485;

RA Baker E.N., Blundell T.L., Cutfield J.F., Cutfield S.M., Dodson E.J.,

RA Dodson G.G., Crowfoot Hodgkin D.M., Hubbard R.E., Isaacs N.W.,

RA Reynolds C.D., Sakabe K., Sakabe N., Vijayan N.M.;

RT "The structure of 2Zn pig insulin crystals at 1.5-A resolution.";
RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 319:369-456(1988).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=92126280; PubMed=1772633;
RA Balschmidt P., Hansen F.B., Dodson E., Dodson G., Korber F.;
RT "Structure of porcine insulin cocrystallized with clupeine Z.";
RL Acta Crystallogr. B 47:975-986(1991).
RN [9]
RP X-RAY CRYSTALLOGRAPHY.
RX MEDLINE=91222450; PubMed=2025410;
RA Badger J., Harris M.R., Reynolds C.D., Evans A.C., Dodson E.J.,
RA Dodson G.G., North A.C.T.;
RT "Structure of the pig insulin dimer in the cubic crystal.";
RL Acta Crystallogr. B 47:127-136(1991).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).
RA Diao J.-S., Wan Z.-L., Chang W.-R., Liang D.-C.;
RT "Structure of monomeric porcine DesB1-B2 despentapeptide (B26-B30)
insulin at 1.65-A resolution.";
RL Acta Crystallogr. D 53:507-512(1997).
CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
increases cell permeability to monosaccharides, amino acids and
fatty acids. It accelerates glycolysis, the pentose phosphate
cycle, and glycogen synthesis in liver.
CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
disulfide bonds.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the insulin family.
CC -!- DATABASE: NAME=Protein Spotlight;
CC NOTE=Issue 9 of April 2001;
CC WWW="<http://www.expasy.org/spotlight/articles/sptlt009.html>".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF064555; AAC77920.1; ALT_INIT.
DR EMBL; AY044828; AAL69550.1; -.
DR PDB; 3INS; 09-JAN-89.
DR PDB; 4INS; 31-JUL-94.
DR PDB; 6INS; 31-JAN-94.
DR PDB; 7INS; 31-JAN-94.
DR PDB; 9INS; 15-OCT-91.
DR PDB; 1IZA; 15-OCT-91.
DR PDB; 1IZB; 15-OCT-91.
DR PDB; 2TCI; 29-JAN-96.
DR PDB; 1MPJ; 29-JAN-96.
DR PDB; 3MTH; 29-JAN-96.
DR PDB; 1DEI; 16-JUN-97.
DR PDB; 1SDB; 01-APR-98.
DR PDB; 1WAV; 28-FEB-97.
DR PDB; 1ZEI; 16-FEB-99.

DR PDB; 1ZNI; 28-JAN-98.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PRO0277; INSULINB.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal; 3D-structure.

FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 85 C PEPTIDE.
FT CHAIN 88 108 INSULIN A CHAIN.
FT DISULFID 31 94 INTERCHAIN.
FT DISULFID 43 107 INTERCHAIN.
FT DISULFID 93 98
FT HELIX 26 46
FT STRAND 48 48
FT HELIX 89 94
FT HELIX 100 106
FT STRAND 107 107
SQ SEQUENCE 108 AA; 11671 MW; CB4491B429858EBE CRC64;

Query Match 89.5%; Score 263; DB 1; Length 108;
Best Local Similarity 60.7%; Pred. No. 7.9e-26;
Matches 51; Conservative 0; Mismatches 1; Indels 32; Gaps 1;

QY 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 30
| | | | | | | | | | | | | | | | | | | | | | | |
Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKARREAENPQAGAVELGGGLGGLQALALEGPP 84
QY 31 --RGIVEQCCTSICSLYQLENYCN 52
| | | | | | | | | | | | | | | | | | | | | | | |
Db 85 QKRGIVEQCCTSICSLYQLENYCN 108

RESULT 12
INS_RABIT
ID INS_RABIT STANDARD PRT; 110 AA.
AC P01311;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Insulin precursor.
GN INS.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Pancreas;
RX MEDLINE=94179230; PubMed=8132571;
RA Devaskar S.U., Giddings S.J., Rajakumar P.A., Carnaghi L.R.,
RA Menon R.K., Zahm D.S.;
RT "Insulin gene expression and insulin synthesis in mammalian neuronal
RT cells.";
RL J. Biol. Chem. 269:8445-8454(1994).
RN [2]

RP SEQUENCE OF 25-54 AND 90-110.
 RX MEDLINE=66160119; PubMed=5949593;
 RA Smith L.F.;
 RT "Species variation in the amino acid sequence of insulin.";
 RL Am. J. Med. 40:662-666(1966).
 RN [3]
 RP SEQUENCE OF 56-110 FROM N.A.
 RA Giddings S.J., Carnaghi L.R., Devaskar S.U.;
 RL Submitted (APR-1991) to the EMBL/GenBank/DDBJ databases.
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
 increases cell permeability to monosaccharides, amino acids and
 fatty acids. It accelerates glycolysis, the pentose phosphate
 cycle, and glycogen synthesis in liver.
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
 disulfide bonds.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to the insulin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U03610; AAA19033.1; -.
 DR EMBL; M61153; AAA17540.1; -.
 DR PIR; A53438; INRB.
 DR HSSP; P01308; 1TYM.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; ILGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 54 INSULIN B CHAIN.
 FT PROPEP 57 87 C PEPTIDE.
 FT CHAIN 90 110 INSULIN A CHAIN.
 FT DISULFID 31 96 INTERCHAIN.
 FT DISULFID 43 109 INTERCHAIN.
 FT DISULFID 95 100
 FT CONFLICT 83 83 E -> Y (IN REF. 3).
 SQ SEQUENCE 110 AA; 11838 MW; 82D2975B85D77FA8 CRC64;

 Query Match 89.5%; Score 263; DB 1; Length 110;
 Best Local Similarity 59.3%; Pred. No. 8e-26;
 Matches 51; Conservative 1; Mismatches 0; Indels 34; Gaps 1;

 Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 30
 |||||||||||||||||:
 Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKSREVEELQVGQAE^{LGGGPAGGLQPSALEL} 84

 Qy 31 ----RGIVEQCCTSICSLYQLENYCN 52
 |||||||||:
 Db 85 ALQKRGIVEQCCTSICSLYQLENYCN 110

RESULT 13

INS_SPETR

ID INS_SPETR STANDARD; PRT; 110 AA.

AC Q91XI3;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Insulin precursor.

GN INS.

OS Spermophilus tridecemlineatus (Thirteen-lined ground squirrel).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;

OC Spermophilus.

OX NCBI_TaxID=43179;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Pancreas;

RA Tredrea M.M., Buck M.J., Guhaniyogi J., Squire T.L., Andrews M.T.;

RT "Regulation of PDK4 expression in a hibernating mammal.";

RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.

CC -!- FUNCTION: Insulin decreases blood glucose concentration. It increases cell permeability to monosaccharides, amino acids and fatty acids. It accelerates glycolysis, the pentose phosphate cycle, and glycogen synthesis in liver.

CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two disulfide bonds.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- SIMILARITY: Belongs to the insulin family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AY038604; AAK72558.1; -.

DR HSSP; P01308; 1LNP.

DR InterPro; IPR004825; Ins/IGF/relax.

DR Pfam; PF00049; Insulin; 1.

DR PRINTS; PR00277; INSULINB.

DR SMART; SM00078; ILGF; 1.

DR PROSITE; PS00262; INSULIN; 1.

KW Insulin family; Hormone; Glucose metabolism; Signal.

FT SIGNAL 1 24 BY SIMILARITY.

FT CHAIN 25 54 INSULIN B CHAIN.

FT PROPEP 57 87 C PEPTIDE.

FT CHAIN 90 110 INSULIN A CHAIN.

FT DISULFID 31 96 INTERCHAIN (BY SIMILARITY).

FT DISULFID 43 109 INTERCHAIN (BY SIMILARITY).

FT DISULFID 95 100 BY SIMILARITY.

SQ SEQUENCE 110 AA; 12004 MW; 4511768D6622BEE5 CRC64;

Query Match 89.5%; Score 263; DB 1; Length 110;

Best Local Similarity 59.3%; Pred. No. 8e-26;
Matches 51; Conservative 1; Mismatches 0; Indels 34; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFYTPKT----- 30
| | | | | | | | | | | | | | | | | | | | | |:
Db 25 FVNQHLCGSHLVEALYLVCGERGFYTPKSRRVEEQQGGQVELGGPGAGLPQPLAEM 84

Qy 31 ----RGIVEQCCTSICSLYQLENYCN 52
| | | | | | | | | | | | | | | |:
Db 85 ALQKRGIVEQCCTSICSLYQLENYCN 110

RESULT 14

INS_FELCA

ID INS_FELCA STANDARD; PRT; 51 AA.
AC P06306;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Insulin.
GN INS.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE.
RX MEDLINE=86214076; PubMed=3518635;
RA Hallden G., Gafvelin G., Mutt V., Joernvall H.;
RT "Characterization of cat insulin.";
RL Arch. Biochem. Biophys. 247:20-27(1986).
CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
increases cell permeability to monosaccharides, amino acids and
fatty acids. It accelerates glycolysis, the pentose phosphate
cycle, and glycogen synthesis in liver.
CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
disulfide bonds.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the insulin family.
DR PIR; A01588; INCT.
DR HSSP; P01317; 1APH.
DR InterPro; IPR004825; Ins/IGF/relax.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5745 MW; 9007B5096A0A7DDD CRC64;

Query Match 89.3%; Score 262.5; DB 1; Length 51;

Best Local Similarity 90.4%; Pred. No. 4.3e-26;
Matches 47; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 1 FVNQHLCGSHLVEALYLVCGERGFFTPKTRGIVEQCCTSICSLYQLENYCN 52
| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 FVNQHLCGSHLVEALYLVCGERGFFTPKA-GIVEQCCASVCSLYQLEHYCN 51

RESULT 15
INS_CANFA
ID INS_CANFA STANDARD; PRT; 110 AA.
AC P01321;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Insulin precursor.
GN INS.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83109071; PubMed=6296142;
RA Kwok S.C.M., Chan S.J., Steiner D.F.;
RT "Cloning and nucleotide sequence analysis of the dog insulin gene.
RT Coded amino acid sequence of canine preproinsulin predicts an
RT additional C-peptide fragment.";
RL J. Biol. Chem. 258:2357-2363(1983).
RN [2]
RP SEQUENCE OF 25-54 AND 90-110.
RX MEDLINE=66160119; PubMed=5949593;
RA Smith L.F.;
RT "Species variation in the amino acid sequence of insulin.";
RL Am. J. Med. 40:662-666(1966).
CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
CC increases cell permeability to monosaccharides, amino acids and
CC fatty acids. It accelerates glycolysis, the pentose phosphate
CC cycle, and glycogen synthesis in liver.
CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
CC disulfide bonds.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the insulin family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; V00179; CAA23475.1; -.
DR PIR; A92413; IPDG.
DR HSPP; P01317; 1APH.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; ILGF; 1.

DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 87 C PEPTIDE.
FT CHAIN 90 110 INSULIN A CHAIN.
FT DISULFID 31 96 INTERCHAIN.
FT DISULFID 43 109 INTERCHAIN.
FT DISULFID 95 100
SQ SEQUENCE 110 AA; 12190 MW; A574791864A4FB98 CRC64;

Query Match 89.1%; Score 262; DB 1; Length 110;
Best Local Similarity 59.3%; Pred. No. 1.1e-25;
Matches 51; Conservative 0; Mismatches 1; Indels 34; Gaps 1;

Qy 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 30
| | | | | | | | | | | | | | | | | | | | | | | |
Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKARREVEDLQVRDVELAGAPGEGLQPLALEG 84

Qy 31 ----RGIVEQCCTSICSLYQLENYCN 52
| | | | | | | | | | | | | | | | | | | | | | | |
Db 85 ALQKRGIVEQCCTSICSLYQLENYCN 110

Search completed: July 15, 2004, 16:36:26
Job time : 4.58955 secs